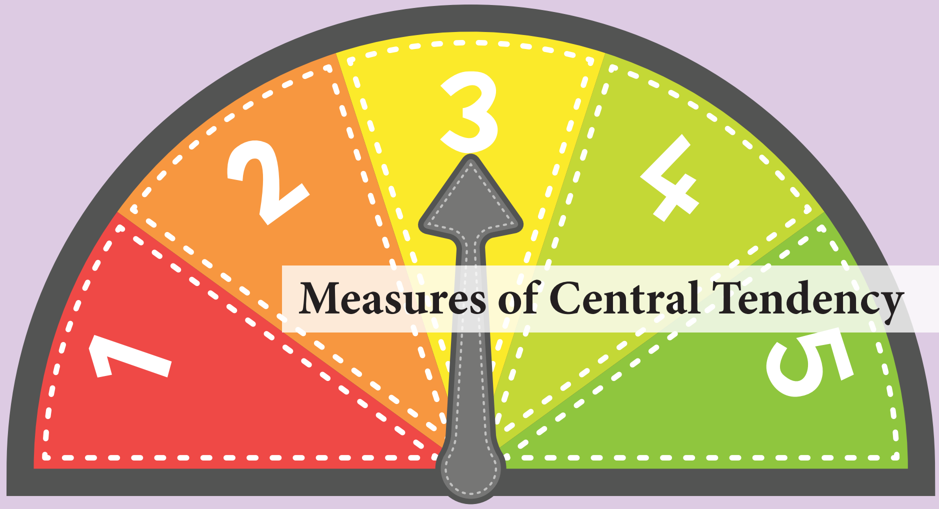
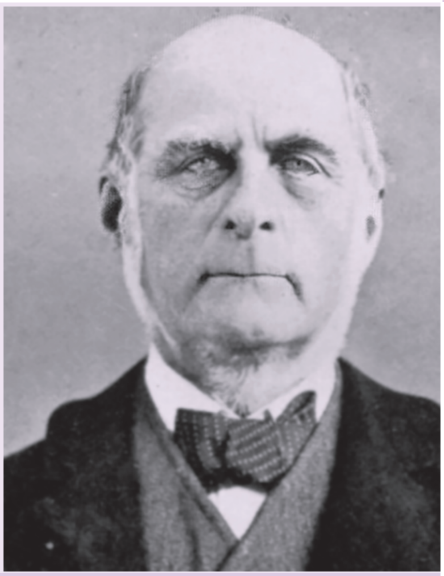
Chapter 5

Measures of Central Tendency





Sir Francis Galton

(16 Feb, 1822- 14 Jan, 1911)

Sir Francis Galton, born in England was a statistician, sociologist, psychologist anthropologist, eugenicist, tropical explorer, geographer, inventor, meteorologist, and psychometrician. Galton produced over 340 papers and several books. He also created the statistical concept of correlation and regression. He was the first to apply statistical methods to the study of human differences and inheritance of intelligence, and introduced the use of questionnaires and surveys for collecting data on human communities.

As an initiator of scientific meteorology, he devised the first weather map, proposed a theory of anticyclones, and was the first to establish a complete record of short-term climatic phenomena on a European scale. He also invented the Galton Whistle for testing differential hearing ability.

'The inherent inability of the human mind to grasp in it's entirely a large body of numerical data compels us to see relatively few constants that will adequately describe the data'.

\- Prof. R. A. Fisher

Learning Objectives

- Knows the average as the representation of the entire group
- Calculates the mathematical averages and the positional averages
- Computes quartiles, Deciles, Percentiles and interprets
- Understands the relationships among the averages and stating their uses.

Introduction

Human mind is incapable of remembering the entire mass of unwieldy data.Having learnt the methods of collection and presentation of data, one has to condense the data to get representative numbers to study the characteristics of data. The characteristics of the data set is explored with some numerical measures namely measures of central tendency, measures of dispersion, measures ofskewness, and measures of kurtosis. This unit focuses on "Measure of central tendency". The measures of central tendency are also called "the averages".

In practical situations one need to have a single value to represent each variable in the whole set of data. Because, the values of the variable are not equal, however there is a general tendency of such observations to cluster around a particular level. In this situation it may be preferable to characterize each group of observations by a single value such that all other values clustered around it. That is why such measure is called the measure of central tendency of that group. A measure of central tendency is a representative value of the entire group of data. It describes the characteristic of the entire mass of data. It reduces the complexity of data and makes them amenable for the application of mathematical techniques involved in analysis and interpretation of data.

5.1 Definition of Measures of Central Tendency

Various statisticians have defined the word average differently. Some of the important definitions are given below:

"Average is an attempt to find one single figure to describe whole of figure"

– Clark and Sekkade

"Average is a value which is typical or representative of a set of data"

– Murray R. Speigal.

"The average is sometimes described as number which is typical of the whole group"

– Leabo.

It is clear from the above definitions that average is a typical value of the entire data and is a measures of central tendency.

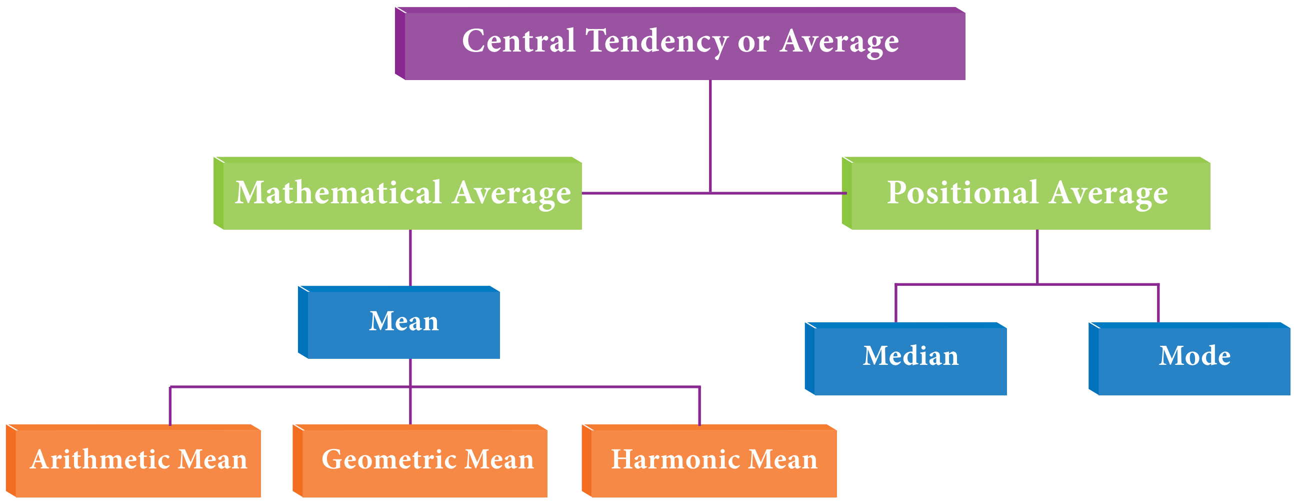
5.2 Characteristics for a good statistical average

The following properties should be possessed by an ideal average.

- It should be well defined so that a unique answer can be obtained.
- It should be easy to understand, calculate and interpret.
- It should be based on all the observations of the data.

- It should be amenable for further mathematical calculations.
- It should be least affected by the fluctuations of the sampling.
- It should not be unduly affected by the extreme values.

5.3 Various measures of central tendency



5.3.1 Arithmetic Mean

(a) To find A.M. for Raw data

For a raw data, the arithmetic mean of a series of numbers is sum of all observations divided by the number of observations in the series. Thus if $x_1, x_2, \ldots, x_n$ represent the values of n observations, then arithmetic mean (A.M.) for n observations is: (direct method)

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

There are two methods for computing the A.M :

(i) Direct method (ii) Short cut method.

NOTE

Normal health parameters such as blood pressure, pulse rate, blood cell count, BMI, blood sugar level etc., are calculated averages of people in a particular region and always vary among individuals.

Example 5.1

The following data represent the number of books issued in a school library on selected from 7 different days 7, 9, 12, 15, 5, 4, 11 find the mean number of books.

Solution:

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

$$\bar{x} = \frac{7+9+12+15+5+4+11}{7}$$

$$= \frac{63}{7} = 9$$

Hence the mean of the number of books is 9

(ii) Short-cut Method to find A.M.

Under this method an assumed mean or an arbitrary value (denoted by A) is used as the basis of calculation of deviations (d_i) from individual values. That is if $d_i = x_i - A$ then

$$\bar{x} = A + \frac{\sum_{i=1}^{n} d_i}{n}$$

Example 5.2

A student's marks in 5 subjects are 75, 68, 80, 92, 56. Find the average of his marks.

Solution:

Let us take the assumed mean, A = 68

x_i	$d_i = x_i - 68$
75	7
68	0
80	12
56	-12
92	24
Total	**31**

$$\bar{x} = A + \frac{\sum_{i=1}^{n} d_i}{n}$$

$$= 68 + \frac{31}{5}$$

$$= 68+6.2 = 74.2$$

The arithmetic mean of average marks is 74.2

(b) To find A.M. for Discrete Grouped data

If $x_1, x_2, ..., x_n$ are discrete values with the corresponding frequencies $f_1, f_2, \ldots, f_n$. Then the mean for discrete grouped data is defined as (direct method)

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

In the short cut method the formula is modified as

$$\bar{x} = A + \frac{\sum_{i=1}^{n} f_i d_i}{N} \quad \text{where} \quad d_i = x_i - A$$

Example 5.3

A proof reads through 73 pages manuscript The number of mistakes found on each of the pages are summarized in the table below Determine the mean number of mistakes found per page

No of mistakes	1	2	3	4	5	6	7
No of pages	5	9	12	17	14	10	6

Solution:

(i) Direct Method

x_i	f_i	$f_i x_i$
1	5	5
2	9	18
3	12	36
4	17	68
5	14	70
6	10	60
7	6	42
Total	N=73	299

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

$$= \frac{299}{73}$$

$$= 4.09$$

The mean number of mistakes is 4.09

(ii) Short-cut Method

x_i	f_i	$d_i = x_i - A$	$f_i d_i$
1	5	-3	-15
2	9	-2	-18

3	12	-1	-12
4	17	0	0
5	14	1	14
6	10	2	20
7	6	3	18
	$\Sigma f_i=73$		$\Sigma f_i d_i=7$

$$\bar{x} = A + \frac{\sum_{i=1}^{n} f_i d_i}{N}$$

$$= 4 + \frac{7}{73}$$

$$= 4.09$$

The mean number of mistakes = 4.09

(c) Mean for Continuous Grouped data:

For the computation of A.M for the continuous grouped data, we can use direct method or short cut method.

(i) Direct Method:

The formula is

$$\bar{x} = \frac{\sum_{i=1}^{n} f_i x_i}{N},$$ x_i is the midpoint of the class interval

(ii) Short cut method

$$\bar{x} = A + \frac{\sum_{i=1}^{n} f_i d_i}{N} \times C$$

$$d = \frac{x_i - A}{c}$$

where A - any arbitrary value

c - width of the class interval

x_i is the midpoint of the class interval.

Example 5.4

The following the distribution of persons according to different income groups

Income (in ₹1000)	0 – 8	8 – 16	16 – 24	24 – 32	32 – 40	40 – 48
No of persons	8	7	16	24	15	7

Find the average income of the persons.

Solution :

Direct Method:

Class	f_i	x_i	$f_i x_i$
0-8	8	4	32
8 – 16	7	12	84
16-24	16	20	320
24-32	24	28	672
32-40	15	36	540
40-48	7	44	308
Total	N =77		1956

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

$$= \frac{1956}{77}$$

$$=25.40$$

Short cut method:

Class	f_i	x_i	$d_i=(x_i-A)/c$	$f_i d_i$
0 – 8	8	4	-3	-24
8 – 16	7	12	-2	-14
16 – 24	16	20	-1	-16
24 – 32	24	28 A	0	0
32 – 40	15	36	1	15
40 – 48	7	44	2	14
Total	**N= 77**			**-25**

$$\overline{x} = A + \frac{\sum_{i=1}^{n} f_i d_i}{N} \times C$$

$$= 28 + \frac{-25}{77} \times 8 \ \ = 25.40$$

Merits

- It is easy to compute and has a unique value.
- It is based on all the observations.
- It is well defined.
- It is least affected by sampling fluctuations.
- It can be used for further statistical analysis.

Limitations

- The mean is unduly affected by the extreme items (outliers).
- It cannot be determined for the qualitative data such as beauty, honesty etc.
- It cannot be located by observations on the graphic method.

When to use?

Arithmetic mean is a best representative of the data if the data set is homogeneous. On the other hand if the data set is heterogeneous the result may be misleading and may not represent the data.

Weighted Arithmetic Mean

The arithmetic mean, as discussed earlier, gives equal importance (or weights) to each observation in the data set. However, there are situations in which values of individual observations in the data set are not of equal importance. Under these circumstances, we may attach, a weight, as an indicator of their importance to each observation value.

Definition

Let $x_1, x_2, ..., x_n$ be the set of n values having weights $w_1, w_2, ..., w_n$ respectively, then the weighted mean is,

$$\overline{x}_w = \frac{w_1 x_1 + w_2 x_2 + ... + w_n x_n}{w_1 + w_2 + ... + w_n} = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i}$$

Uses of weighted arithmetic mean

Weighted arithmetic mean is used in:

- The construction of index numbers.
- Comparison of results of two or more groups where number of items in the groups differs.
- Computation of standardized death and birth rates.

Example 5.5

The weights assigned to different components in an examination or Component Weightage Marks scored

Component	Weightage	Marks scored
Theory	4	60
Practical	3	80
Assignment	1	90
Project	2	75
	10	

Calculate the weighted average score of the student who scored marks as given in the table

Solution:

Component	Marks scored (x_i)	Weightage (w_i)	$w_i x_i$
Theory	60	4	240
Practical	80	3	240
Assignment	90	1	90
Project	75	2	150
Total		10	720

Weighted average, $\bar{x}_w = \dfrac{\sum w_i x_i}{\sum w_i}$

$= 720/10$

$= 72$

NOTE

If the weights are not assigned, then the mean is

$= (60 + 80 + 90 + 75)/4$

$= 76.25$

Combined Mean:

Let $\overline{x}_1$ and $\overline{x}_2$ are the arithmetic mean of two groups (having the same unit of measurement of a variable), based on n1 and n2 observations respectively. Then the combined mean can be calculated using

$$\text{Combined Mean} = \bar{x}_{12} = \frac{n_1\overline{x}1 + n_2\overline{x}2}{n_1 + n_2}$$

Remark : The above result can be extended to any number of groups.

Example 5.6

A class consists of 4 boys and 3 girls. The average marks obtained by the boys and girls are 20 and 30 respectively. Find the class average.

Solution:

$n_1 = 4,\ \overline{x}_1 = 20,\ n_2 = 3,\ \overline{x}_2 = 30$

$$\text{Combined Mean} = \bar{x}_{12} = \frac{n_1\overline{x}1 + n_2\overline{x}2}{n_1 + n_2}$$

$$= \left[\frac{4 \times 20 + 3 \times 30}{4 + 3}\right]$$

$$= \left[\frac{80 + 90}{7} = \frac{170}{7}\right] = 24.3$$

5.3.2 Geometric Mean(GM)

(a) G.M. For Ungrouped data

The Geometric Mean (G.M.) of a set of n observations is the nth root of their product. If $x_1, x_2, \ldots, x_n$ are n observations then

$$\text{G. M.} = \sqrt[n]{x_1, x_2 \ldots x_n} = (x_1 . x_2 \ldots x_n)^{\frac{1}{n}}$$

Taking the nth root of a number is difficult. Thus, the computation is done as under

$$\log \text{G.M.} = \log (x_1 . x_2 \ldots x_n)$$

$$= (\log x_1 + \log x_2 + \ldots + \log x_n)$$

$$= \frac{\sum_{i=l}^{n} \log x_i}{n}$$

$$\text{G.M.} = \text{Antilog} \frac{\sum_{i=l}^{n} \log x_i}{n}$$

Example 5.7

Calculate the geometric mean of the annual percentage growth rate of profits in business corporate from the year 2000 to 2005 is given below

50, 72, 54, 82, 93

Solution:

x_i	50	72	54	82	93	**Total**
log x_i	1.6990	1.8573	1.7324	1.9138	1.9685	**9.1710**

$$\text{G.M.} = \text{Antilog} \frac{\sum_{i=1}^{n} \log x_i}{n}$$

$$= \text{Antilog} \frac{9.1710}{5}$$

$$= \text{Antilog } 1.8342$$

$$\text{G. M.} = 68.26$$

Geometrical mean of annual percentage growth rate of profits is 68.26

Example 5.8

The population in a city increased at the rate of 15% and 25% for two successive years. In the next year it decreased at the rate of 5%. Find the average rate of growt

Solution:

Let us assume that the population is 100

Percentage rise in population	Population at the end of year x_i	logx_i
15	115	2.0607
25	125	2.0969
5	95	1.9777
		6.1353

$$\text{G.M} = \text{Antilog} \frac{\sum_{n}^{n} \log x_i}{n}$$

$$= \text{Antilog} \frac{(6.1353)}{3}$$

$$= \text{Antilog}\ (2.0451)$$

$$= 110.9$$

(b) G.M. For Discrete grouped data

If $x_1, x_2, \ldots\ldots x_n$ are discrete values of the variate x with corresponding frequencies $f_1, f_2, \ldots f_n$. Then geometric mean is defined as

$$\text{G. M.} = \text{Antilog} \frac{\sum_{i=1}^{n} f_i \log x_i}{N} \text{ with usual notations}$$

Example 5.9

Find the G.M for the following data, which gives the defective screws obtained in a factory.

Diameter (cm)	5	15	25	35
Number of defective screws	5	8	3	4

Solution:

x_i	f_i	$\log x_i$	$f_i \log x_i$
5	5	0.6990	3.4950
15	8	1.1761	9.4088
25	3	1.3979	4.1937
35	4	1.5441	6.1764
	N=20		23.2739

$$\text{G.M} = \text{Antilog}$$

$$= \text{Antilog} \frac{\sum_{i=1}^{n} f_i \log x_i}{N}$$

$$= \text{Antilog} \frac{23.2739}{20}$$

$$= \text{Antilog } 1.1637$$

$$\text{G.M} = 14.58$$

(c) G.M. for Continuous grouped data

Let x_i be the mid point of the class interval

$$\text{G. M.} = \text{Antilog}\left[\frac{\sum_{i=l}^{n} f_i \log x_i}{N}\right]$$

Example 5.10

The following is the distribution of marks obtained by 109 students in a subject in an institution. Find the Geometric mean.

Marks	4-8	8-12	12-16	16-20	20-24	24-28	28-32	32-36	36-40
No. of Students	6	10	18	30	15	12	10	6	2

Solution:

Marks	Mid point (x_i)	f_i	$\log x_i$	$f_i \log x_i$
4-8	6	6	0.7782	4.6692
8-12	10	10	1.0000	10.0000
12-16	14	18	1.1461	20.6298
16-20	18	30	1.2553	37.6590
20-24	22	15	1.3424	20.1360
24-28	26	12	1.4150	16.980
28-32	30	10	1.4771	14.7710
32-36	34	6	1.5315	9.1890
36-40	38	2	1.5798	3.1596
Total		**N =109**		**137.1936**

$$\text{G.M.} = \text{Antilog}\left[\frac{\sum_{i=l}^{n} f_i \log x_i}{N}\right]$$

$$= \text{Antilog}\left[\frac{137.1936}{109}\right] = \text{Antilog}\,[1.2587]$$

$$\text{G. M.} = 18.14$$

Geometric mean marks of 109 students in a subject is 18.14

Merits of Geometric Mean:

- It is based on all the observations
- It is rigidly defined
- It is capable of further algebraic treatment
- It is less affected by the extreme values
- It is suitable for averaging ratios, percentages and rates.

Limitations of Geometric Mean:

- It is difficult to understand
- The geometric mean cannot be computed if any item in the series is negative or zero.
- The GM may not be the actual value of the series
- It brings out the property of the ratio of the change and not the absolute difference of change as the case in arithmetic mean.

5.3.3 Harmonic Mean (H.M.)

Harmonic Mean is defined as the reciprocal of the arithmetic mean of reciprocals of the observations.

(a) H.M. for Ungrouped data

Let $x_1, x_2, ..., x_n$ be the n observations then the harmonic mean is defined as

$$\text{H. M.} = \frac{n}{\sum_{i=l}^{n}\left(\frac{1}{x_i}\right)}$$

Example 5.11

A man travels from Jaipur to Agra by a car and takes 4 hours to cover the whole distance. In the first hour he travels at a speed of 50 km/hr, in the second hour his speed is 65 km/hr, in third hour his speed is 80 km/hr and in the fourth hour he travels at the speed of 55 km/hr. Find the average speed of the motorist.

Solution:

x	50	65	80	55	**Total**
$1/x$	0.0200	0.0154	0.0125	0.0182	**0.0661**

$$\text{H. M.} = \frac{n}{\sum\left(\frac{1}{x_i}\right)}$$

$$= \frac{4}{0.0661} = 60.5 \text{ km/hr}$$

Average speed of the motorist is 60.5km/hr

(b) H.M. for Discrete Grouped data:

For a frequency distribution

$$\text{H. M.} = \frac{N}{\sum_{i=l}^{n} f_i\left(\frac{1}{x_i}\right)}$$

Example 5.12

The following data is obtained from the survey. Compute H.M

Speed of the car	130	135	140	145	150
No of cars	3	4	8	9	2

Solution:

x_i	f_i	$\frac{f_i}{x_i}$
130	3	0.0296
135	4	0.0091
140	8	0.0571
145	9	0.0621
150	2	0.0133
Total	N = 26	0.1852

$$\text{H. M.} = \frac{N}{\sum_{i=l}^{n} f_i\left(\frac{1}{x_i}\right)}$$

$$= \frac{26}{0.1852}$$

$$\text{H.M} = 140.39$$

(c) H.M. for Continuous data:

The Harmonic mean H.M. = $\dfrac{N}{\sum_{i=1}^{n} f_i\left(\dfrac{1}{x_i}\right)}$

Where x_i is the mid-point of the class interval

Example 5.13

Find the harmonic mean of the following distribution of data

Dividend yield (percent)	2 – 6	6 – 10	10 – 14
No. of companies	10	12	18

Solution:

Class Intervals	Mid-value (x_i)	No. of companies (f_i)	Reciprocal ($1/x_i$)	f_i ($1/x_i$)
2 – 6	4	10	¼	2.5
6 – 10	8	12	1/8	1.5
10 – 14	12	18	1/12	1.5
Total		N = 40		5.5

The harmonic mean is H.M. = $\dfrac{N}{\sum_{i=1}^{n} f_i\left(\dfrac{1}{x_i}\right)} = \dfrac{40}{5.5} = 7.27$

Merits of H.M:

- It is rigidly defined
- It is based on all the observations of the series
- It is suitable in case of series having wide dispersion
- It is suitable for further mathematical treatment
- It gives less weight to large items and more weight to small items

Limitations of H.M:

- It is difficult to calculate and is not understandable
- All the values must be available for computation
- It is not popular due to its complex calculation.
- It is usually a value which does not exist in series

When to use?

Harmonic mean is used to calculate the average value when the values are expressed as value/unit. Since the speed is expressed as km/hour, harmonic mean is used for the calculation of average speed.

Relationship among the averages:

In any distribution when the original items are different the A.M., G.M. and H.M would also differ and will be in the following order:

$$\textbf{A.M.} \geq \textbf{G.M} \geq \textbf{H.M}$$

5.3.4 Median

Median is the value of the variable which divides the whole set of data into two equal parts. It is the value such that in a set of observations, 50% observations are above and 50% observations are below it. Hence the median is a positional average.

(a) Median for Ungrouped or Raw data:

In this case, the data is arranged in either ascending or descending order of magnitude.

(i) If the number of observations n is an odd number, then the median is represented by the numerical value of x, corresponds to the positioning point of $\frac{n+1}{2}$ in ordered observations. That is,

$$\text{Median} = \text{value of } \left(\frac{n+1}{2}\right)^{th} \text{ observation in the data array}$$

(ii) If the number of observations n is an even number, then the median is defined as the arithmetic mean of the middle values in the array That is,

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

Example 5.14

The number of rooms in the seven five stars hotel in Chennai city is 71, 30, 61, 59, 31, 40 and 29. Find the median number of rooms

Solution:

Arrange the data in ascending order 29, 30, 31, 40, 59, 61, 71

$$n = 7 \text{ (odd)}$$

$$\text{Median} = \frac{7+1}{2} = 4^{\text{th}} \text{ positional value}$$

$$\text{Median} = 40 \text{ rooms}$$

Example 5.15

The export of agricultural product in million dollars from a country during eight quarters in 1974 and 1975 was recorded as 29.7, 16.6, 2.3, 14.1, 36.6, 18.7, 3.5, 21.3

Find the median of the given set of values

Solution:

We arrange the data in descending order

36.6, 29.7, 21.3, 18.7, 16.6, 14.1, 3.5, 2.3

$$n = 8 \text{ (even)}$$

$$\text{Median} = \frac{4^{\text{th}}\text{item} + 5^{\text{th}} \text{ item}}{2}$$

$$= \frac{18.7 + 16.6}{2}$$

$$= 17.65 \text{ million dollars}$$

Cumulative Frequency

In a grouped distribution, values are associated with frequencies. The cumulative frequencies are calculated to know the total number of items above or below a certain limit.This is obtained by adding the frequencies successively up to the required level. This cumulative frequencies are useful to calculate median, quartiles, deciles and percentiles.

(b) Median for Discrete grouped data

We can find median using following steps

(i) Calculate the cumulative frequencies

(ii) Find $\frac{N+1}{2}$, Where $N = \sum f$ = total frequencies

(iii) Identify the cumulative frequency just greater than $\frac{N+1}{2}$

(iv) The value of x corresponding to that cumulative frequency $\frac{N+1}{2}$ is the median.

Example 5.16

The following data are the weights of students in a class. Find the median weights of the students

Weight(kg)	10	20	30	40	50	60	70
Number of Students	4	7	12	15	13	5	4

Solution:

Weight (kg) x	Frequency f	Cumulative Frequency $c.f$
10	4	4
20	7	11
30	12	23
40	15	38
50	13	51
60	5	56
70	4	60
Total	N = 60	

Here, $\quad N = \sum f = 60$

$$\frac{N+1}{2} = 30.5$$

The cumulative frequency greater than 30.5 is 38.The value of x corresponding to 38 is 40. The median weight of the students is 40 kgs

(c) Median for Continuous grouped data

In this case, the data is given in the form of a frequency table with class-interval etc., The following formula is used to calculate the median.

$$\text{Median} = l + \frac{\frac{N}{2} - m}{f} \times c$$

Where

l = Lower limit of the median class

N = Total Numbers of frequencies

f = Frequency of the median class

NOTE

If the class intervals are given in inclusive type, convert them into exclusive type and call it as true class interval and consider the lower limit in it.

m = Cumulative frequency of the class preceding the median class

c = the class interval of the median class.

From the formula, it is clear that one has to find the median class first. Median class is, that class which correspond to the cumulative frequency just greater than $\frac{N}{2}$.

Example 5.17

The following data attained from a garden records of certain period Calculate the median weight of the apple

Weight in grams	410 – 420	420 – 430	430 – 440	440 – 450	450 – 460	460 – 470	470 – 480
Number of apples	14	20	42	54	45	18	7

Solution:

Weight in grams	Number of apples	Cumulative Frequency
410 – 420	14	14
420 – 430	20	34
430 – 440	42	76
440 – 450	54	130
450 – 460	45	175
460 – 470	18	193
470 – 480	7	200
Total	**N = 200**	

$$\frac{N}{2} = \frac{200}{2} = 100.$$

Median class is 440 – 450

$$\text{Median } = l + \frac{\frac{N}{2} - m}{f} \times c$$

$$l = 440, \quad \frac{N}{2} = 100, \quad m = 76, \quad f = 54, \quad c = 10$$

$$\text{Median } = 440 + \frac{100 - 76}{54} \times 10$$

$$= 440 + \frac{24}{54} \times 10 = 440 + 4.44 = 444.44$$

The median weight of the apple is 444.44 grams

Example 5.18

The following table shows age distribution of persons in a particular region:

Age (years)	No. of persons (in thousands)
Below 10	2
Below20	5
Below30	9
Below 40	12
Below 50	14
Below 60	15
Below 70	15.5
Below 80	15.6

Find the median age.

Solution:

We are given upper limit and less than cumulative frequencies. First find the class-intervals and the frequencies. Since the values are increasing by 10, hence the width of the class interval is equal to 10.

Age groups	No. of persons (in thousands) f	cf
0 – 10	2	2
10 – 20	3	5
20 – 30	4	9
30 – 40	3	12
40 – 50	2	14
50 – 60	1	15
60 – 70	0.5	15.5
70 – 80	0.1	15.6
Total	**N = 15.6**	

$$\left(\frac{N}{2}\right) = \frac{15.6}{2} = 7.8$$

Median lies in the 20 – 30 age group

$$\text{Median } = l + \frac{\frac{N}{2} - m}{f} \times c$$

$$= 20 + \frac{7.8 - 5}{4} \times 10$$

$$\text{Median } = 27 \text{ years}$$

Example 5.19

The following is the marks obtained by 140 students in a college. Find the median marks

Marks	Number of students
10-19	7
20-29	15
30-39	18
40-49	25
50-59	30
60-69	20
70-79	16
80-89	7
90-99	2

NOTE

In this problem the class intervals are given in inclusive type, convert them into exclusive type and call it as true class interval.

Solution:

Class boundaries	f	Cf
9.5 -19.5	7	7
19.5-29.5	15	22
29.5- 39.5	18	40
39.5-49.5	25	65
49.5-59.5	30	95
59.5-69.5	20	115
69.5-79.5	16	131
79.5-89.5	7	138
89.5-99.5	2	140
Total	**N =140**	

$$\text{Median} = l + \left(\frac{\frac{N}{2} - m}{f} \right) \times c$$

$$\frac{N}{2} = \frac{140}{2} = 70$$

Here $l = 49.5$, $f = 30$, $m = 65$, $c = 10$

$$\text{Median} = 49.5 + \left(\frac{70 - 65}{30} \right) \times 10$$

$$= 49.5 + 1.67$$

$$= 51.17$$

Graphical method for Location of median

Median can be located with the help of the cumulative frequency curve or 'ogive'. The procedure for locating median in a grouped data is as follows:

Step 1 : The class intervals, are represented on the horizontal axis (x-axis)

Step 2 : The cumulative frequency corresponding to different classes is calculated. These cumulative frequencies are plotted on the vertical axis (y-axis) against the upper limit of the respective class interval

Step 3 : The curve obtained by joining the points by means of freehand is called the *'less than ogive'*.

Step 4 : A horizontal straight line is drawn from the value $\frac{N}{2}$ or $\frac{N+1}{2}$ on the y-axis parallel to x- axis to meet the ogive. (depending on N is odd or even)

Step 5 : From the point of intersection, draw a line, perpendicular to the horizontal axis which meet the x axis at m say.

Step 6 : The value m at x axis gives the value of the median.

Remarks:

(i) Similarly 'more than' ogives, can be drawn by plotting more than cumulative frequencies against lower limit of the class. A horizontal straight line is drawn from the value $\frac{N}{2}$ or $\frac{N+1}{2}$ on the y-axis parallel to x-axis to meet the ogive. A line is drawn perpendicular to x-axis meets the point at m, say, the X coordinate of m gives the value of the median.

(depending on N is odd or even)

(ii) When the two ogive curves are drawn on the same graph, a line is drawn perpendicular to x-axis from the point of intersection, meets the point at m, say. The x coordinate m gives the value of the median.

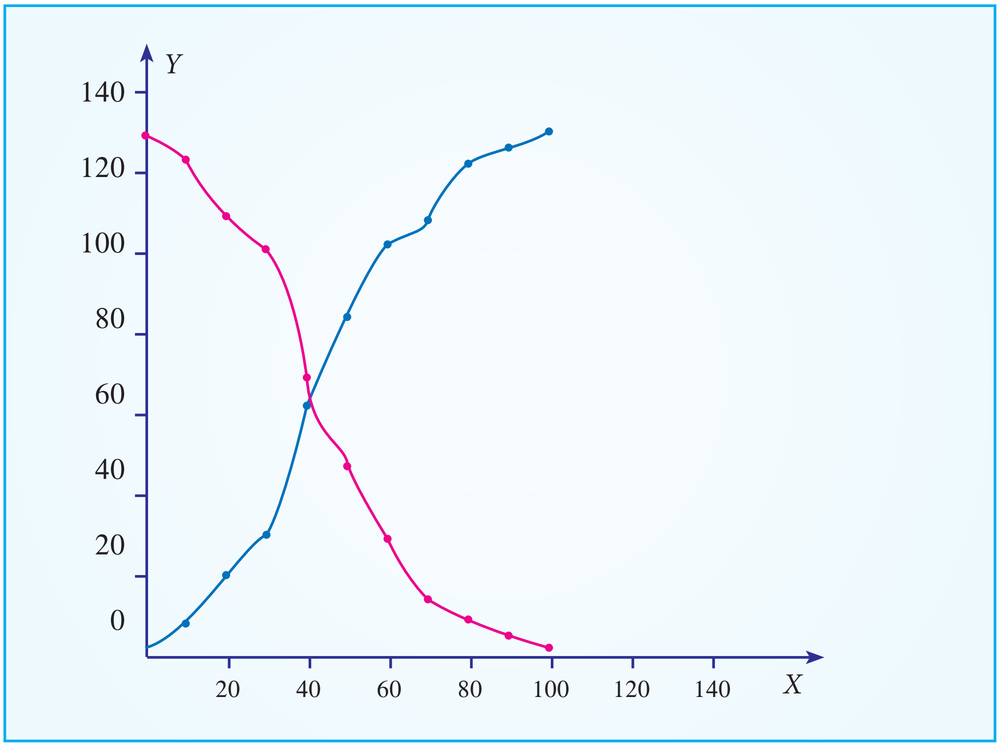
Example 5.20

Draw ogive curves for the following frequency distribution and determine the median.

Age groups	No. of people
0 – 10	6
10 – 20	12
20 – 30	10
30 – 40	32
40 – 50	22
50 – 60	18
60 – 70	15
70 – 80	5
80 – 90	4
90 – 100	3

Solution:

Class boundary	Cumulative Frequency	
	Less than	More than
0	0	127
10	6	121
20	18	109
30	28	99
40	60	67
50	82	45
60	100	27
70	115	12
80	120	7
90	124	3
100	127	0



The median value from the graph is 42

Merits

- It is easy to compute. It can be calculated by mere inspection and by the graphical method
- It is not affected by extreme values.
- It can be easily located even if the class intervals in the series are unequal

Limitations

- It is not amenable to further algebraic treatment
- It is a positional average and is based on the middle item
- It does not take into account the actual values of the items in the series

5.3.5 Mode

According to Croxton and Cowden, 'The mode of a distribution is the value at the point around which the items tend to be most heavily concentrated'.

In a busy road, where we take a survey on the vehicle - traffic on the road at a place at a particular period of time, we observe the number of two wheelers is more than cars, buses and other vehicles. Because of the higher frequency, we say that the modal value of this survey is ‘two wheelers’

Mode is defined as the value which occurs most frequently in a data set. The mode obtained may be two or more in frequency distribution.

Computation of mode:

(a) For Ungrouped or Raw Data:

The mode is defined as the value which occurs frequently in a data set

Example 5.21

The following are the marks scored by 20 students in the class. Find the mode
90, 70, 50, 30, 40, 86, 65, 73, 68, 90, 90, 10, 73, 25, 35, 88, 67, 80, 74, 46

Solution:

Since the marks 90 occurs the maximum number of times, three times compared with the other numbers, mode is 90.

Example 5.22

A doctor who checked 9 patients’ sugar level is given below. Find the mode value of the sugar levels. 80, 112, 110, 115, 124, 130, 100, 90, 150, 180

Solution:

Since each values occurs only once, there is no mode.

Example 5.23

Compute mode value for the following observations.

2, 7, 10, 12, 10, 19, 2, 11, 3, 12

Solution:

Here, the observations 10 and 12 occurs twice in the data set, the modes are 10 and 12.

For discrete frequency distribution, mode is the value of the variable corresponding to the maximum frequency.

NOTE

It is clear that mode may not exist or mode may not be unique.

Example 5.24

Calculate the mode from the following data

Days of Confinement	6	7	8	9	10
Number of patients	4	6	7	5	3

Solution:

Here, 7 is the maximum frequency, hence the value of x corresponding to 7 is 8. Therefore 8 is the mode.

(b) Mode for Continuous data:

The mode or modal value of the distribution is that value of the variate for which the frequency is maximum. It is the value around which the items or observations tend to be most heavily concentrated. The mode is computed by the formula.

$$\text{Mode} = l + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times c$$

Modal class is the class which has maximum frequency.

f_1 = frequency of the modal class

f_0 = frequency of the class preceding the modal class

f_2 = frequency of the class succeeding the modal class

c = width of the class limits

Remarks

(i) If $(2f_1 - f_0 - f_2)$ comes out to be zero, then mode is obtained by the following formula taking absolute differences $M_0 = l + \left(\frac{(f_1 - f_0)}{|f_1 - f_0| + |f_1 - f_2|} \times C\right)$

(ii) If mode lies in the first class interval, then f_0 is taken as zero.

(iii) The computation of mode poses problem when the modal value lies in the open-ended class.

Example 5.25

The following data relates to the daily income of families in an urban area. Find the modal income of the families.

Income (₹)	0-100	100-200	200-300	300-400	400-500	500-600	600-700
No.of persons	5	7	12	18	16	10	5

Solution:

Income (₹)	No.of persons (f)	
0-100	5	
100-200	7	
200-300	12	f_0
300-400	18	f_1
400-500	16	f_2
500-600	10	
600-700	5	

$$\text{Mode} = l + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times C$$

The highest frequency is 18, the modal class is 300-400

Here, $l = 300$, $f_0 = 12$, $f_1 = 18$, $f_2 = 16$,

$$\text{Mode} = 300 + \frac{18 - 12}{2 \times 18 - 12 - 16} \times 100$$

$$= 300 + \frac{6}{36 - 28} \times 100$$

$$= 300 + \frac{6}{8} \times 100$$

$$= 300 + \frac{600}{8} = 300 + 75 = 375$$

The modal income of the families is 375.

Determination of Modal class:

For a frequency distribution modal class corresponds to the class with maximum frequency. But in any one of the following cases that is not easily possible.

(i) If the maximum frequency is repeated.

(ii) If the maximum frequency occurs in the beginning or at the end of the distribution

(iii) If there are irregularities in the distribution, the modal class is determined by the method of grouping.

Steps for preparing Analysis table:

We prepare a grouping table with 6 columns

(i) In column I, we write down the given frequencies.

(ii) Column II is obtained by combining the frequencies two by two.

(iii) Leave the Ist frequency and combine the remaining frequencies two by two and write in column III

(iv) Column IV is obtained by combining the frequencies three by three.

(v) Leave the Ist frequency and combine the remaining frequencies three by three and write in column V

NOTE

In discrete or continuous series, an error of judgement is possible by finding y inspection. In these cases, where the difference between the maximum frequency and the frequency preceeding or succeeding is very small and the items are heavily concentrated on either side, under such circumstance the value of mode is determined by preparing grouping table and through analysis table.

(vi) Leave the Ist and 2nd frequencies and combine the remaining frequencies three by three and write in column VI

Mark the highest frequency in each column. Then form an analysis table to find the modal class. After finding the modal class use the formula to calculate the modal value.

Example 5.26

Calculate mode for the following frequency distribution:

Size	0-5	5-10	10-15	15-20	20-25	25-30	30-35	35-40
Frequency	9	12	15	16	17	15	10	13

Solution:

class	f	2	3	4	5	6
0-5	9	21		36		
5-10	12		27			
10-15	15	31			43	
15-20	16					48
20-25	17	32	33	48		
25-30	15					
30-35	10	23	25		42	38
35-40	13					

Analysis Table:

Columns	0-5	5-10	10-15	15-20	20-25	25-30	30-35	35-40
1					1			
2					1	1		
3				1	1			
4				1	1	1		
5		1	1	1				
6			1	1	1			
Total		1	2	4	5	2		

The maximum occurred corresponding to 20-25, and hence it is the modal class.

$$\text{Mode} = l + \frac{f_1\ f_0}{2f_1 - f_0 - f_2} \times C$$

Here, $l = 20$, $f_0 = 16$, $f_1 = 17$, $f_2 = 15$

$$= 20 + \frac{17-16}{2\times 17-16-15} \times C$$

$$= 20 + \frac{1}{34-31} \times 5$$

$$= 20 + \frac{5}{3} = 20 + 1.67 = 21.67$$

Mode = 21.67

(d) Graphical Location of Mode

The following are the steps to locate mode by graph

(i) Draw a histogram of the given distribution.

(ii) Join the rectangle corner of the highest rectangle (modal class rectangle) by a straight line to the top right corner of the preceding rectangle. Similarly the top left corner of the highest rectangle is joined to the top left corner of the rectangle on the right.

(iii) From the point of intersection of these two diagonal lines, draw a perpendicular line to the x –axis which meets at M.

(iv) The value of x coordinate of M is the mode.

Example 5.27

Locate the modal value graphically for the following frequency distribution

Class Interval	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60
Frequency	5	8	12	7	5	3

Solution:

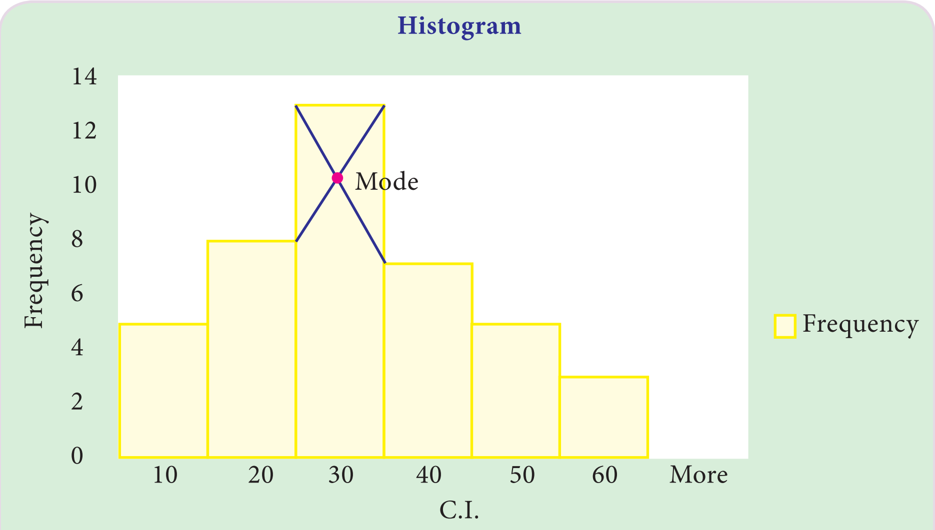


Fig.

Merits of Mode:

- It is comparatively easy to understand.
- It can be found graphically.
- It is easy to locate in some cases by inspection.
- It is not affected by extreme values.
- It is the simplest descriptive measure of average.

Demerits of Mode:

- It is not suitable for further mathematical treatment.
- It is an unstable measure as it is affected more by sampling fluctuations.
- Mode for the series with unequal class intervals cannot be calculated.
- In a bimodal distribution, there are two modal classes and it is difficult to determine the values of the mode.

5.4 Empirical Relationship among mean, median and mode

A frequency distribution in which the values of arithmetic mean, median and mode coincide is known of symmetrical distribution, when the values of mean, median and mode are not equal the distribution is known as asymmetrical or skewed. In moderately skewed asymmetrical distributions a very important relationship exists among arithmetic mean, median and mode.

Karl Pearson has expressed this relationship as follows

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

Example 5.28

In a moderately asymmetrical frequency distribution, the values of median and arithmetic mean are 72 and 78 respectively; estimate the value of the mode.

Solution:

The value of the mode is estimated by applying the following formula:

$$\text{Mode} = 3\ \text{Median} - 2\ \text{Mean} = 3\ (72) - 2\ (78)$$

$$= 216 - 156 = 60$$

$$\text{Mode} = 60$$

Example 5.29

In a moderately asymmetrical frequency distribution, the values of mean and mode are 52.3 and 60.3 respectively, Find the median value.

Solution:

The value of the median is estimated by applying the formula:

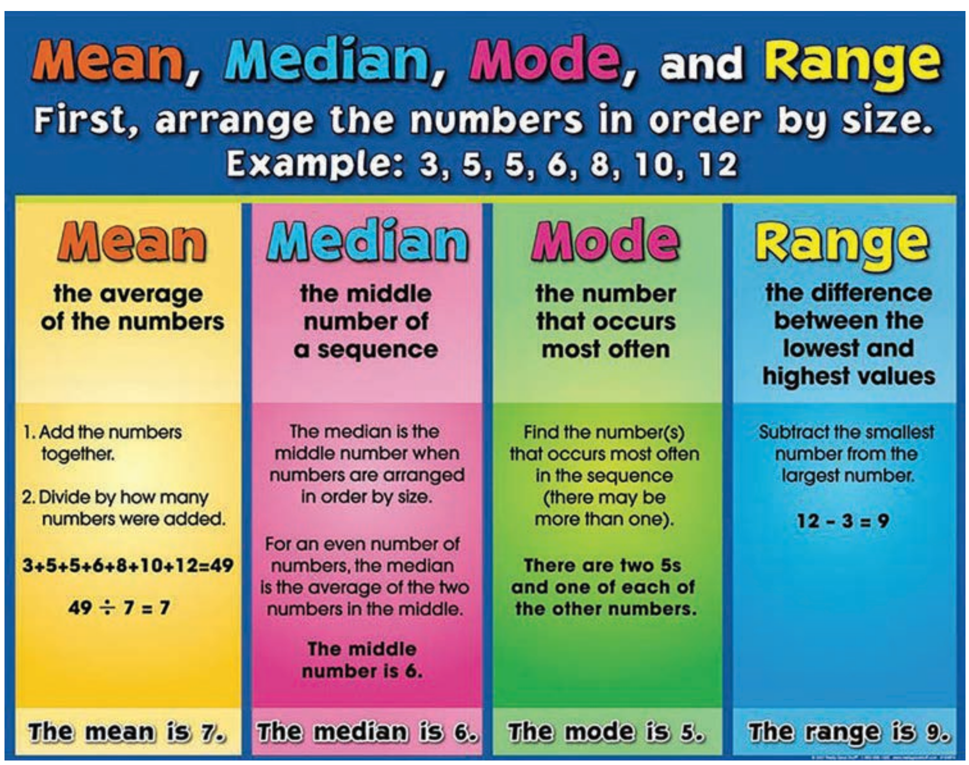
$$\text{Mode} = 3\text{ Median} - 2\text{ Mean}$$

$$60.3 = 3\text{ Median} - 2 \times 52.3$$

$$3\text{ Median} = 60.3 + 2 \times 52.3$$

$$60.3 + 104.6 = 164.9$$

$$\text{Median} = \frac{164.9}{3} = 54.966 = 54.97$$



5.5 Partition Measures

5.5.1 Quartiles

There are three quartiles denoted by Q_1, Q_2 and Q_3 divides the frequency distribution in to four equal parts

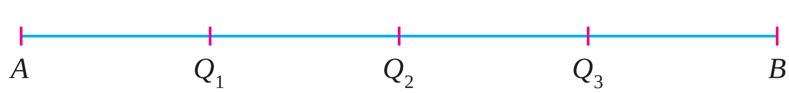


Fig.

That is 25 percent of data will lie below Q_1, 50 percent of data below Q_2 and 75 percent below Q_3. Here Q_2 is called the Median. Quartiles are obtained in almost the same way as median

Quartiles for Raw or Ungrouped data:

If the data set consist of n items and arranged in ascending order then

$$Q_1 = \left(\frac{n+1}{4}\right)^{th} \text{ item}, \quad Q_2 = \left(\frac{n+1}{2}\right)^{th} \text{ item and } \quad Q_3 = 3\left(\frac{n+1}{4}\right)^{th} \text{ item}$$

Example 5.30

Compute Q_1 and Q_3 for the data relating to the marks of 8 students in an examination given below 25, 48, 32, 52, 21, 64, 29, 57

Solution:

$$n = 8$$

Arrange the values in ascending order

21, 25, 29, 32, 48, 52, 57, 64 we have

$$Q_1 = \left(\frac{n+1}{4}\right)^{th} \text{ item}$$

$$= \left(\frac{8+1}{4}\right)^{th} \text{ item}$$

$$= 2.25^{\text{th}} \text{ item}$$

$$= 2^{\text{nd}} \text{ item} + \left(\frac{1}{4}\right) (3^{\text{rd}} \text{ item} - 2^{\text{nd}} \text{ item})$$

$$= 25 + 0.25\,(29 - 25)$$

$$= 25 + 1.0$$

$$Q_1 = 26$$

$$Q_3 = 3\left(\frac{n+1}{4}\right)^{th} \text{ item}$$

$$= 3 \times (2.25)^{th} \text{ item}$$

$$= 6.75^{\text{th}} \text{ item}$$

$$= 6^{\text{th}} \text{ item} + 0.75\,(7^{\text{th}} \text{ item} - 6^{\text{th}} \text{ item})$$

$= 52 + (0.75)\ (57 - 52)$

$= 52 + 3.75$

$Q_3 = 55.75$

Quartiles for Discrete Series (grouped data)

Step 1 : Find cumulative frequencies

Step 2 : Find $\left(\frac{N+1}{4}\right)$

Step 3 : See in the cumulative frequencies, the value just greater than $\left(\frac{N+1}{4}\right)$, the corresponding value of x is Q_1

Step 4 : Find $3\left(\frac{N+1}{4}\right)$

Step 5 : See in the cumulative frequencies, the value just greater than $3\left(\frac{N+1}{4}\right)$ then the corresponding value of x is $Q3$.

Example 5.31

Compute Q_1 and Q_3 for the data relating to age in years of 543 members in a village

Age in years	20	30	40	50	60	70	80
No. of members	3	61	132	153	140	51	3

Solution:

x	f	cf
20	3	3
30	61	64
40	132	196
50	153	349
60	140	489
70	51	540
80	3	543

$$Q_1 = \left(\frac{N+1}{4}\right)^{th} \text{ item}$$

$$= \left(\frac{543+1}{4}\right)^{th} \text{ item}$$

$$= 136^{th} \text{ item}$$

Q_1 = 40 years

$$Q_3 = 3\left(\frac{N+1}{4}\right)^{th} \text{ item}$$

$$= 3\left(\frac{543+1}{4}\right)^{th} \text{ item}$$

$$= 3 \times 136^{th} \text{ item}$$

$$= 408^{th} \text{ item}$$

Q_3 = 60 years

Quartiles for Continuous series (grouped data)

Step 1 : Find Cumulative frequencies

Step 2 : Find $\left(\frac{N}{4}\right)$

Step 3: Q_1 class is the class interval corresponding to the value of the cumulative frequency just greater than $\left(\frac{N}{4}\right)$

Step 4 : Q3 class is the class interval corresponding to the value of the cumulative frequency just greater than 3 $\left(\frac{N}{4}\right)$

$$Q_1 = l_1 + \frac{\frac{N}{4} - m_1}{f_1} \times c_1 \quad \text{and} \quad Q_3 = l_3 + \frac{3\left(\frac{N}{4}\right) - m_3}{f_3} \times C_3$$

where N=Σf = total of all frequency values

l_1 = lower limit of the first quartile class

f_1 = frequency of the first quartile class

c_1 = width of the first quartile class

m_1 = c.f. preceding the first quartile class

l_3 = lower limit of the 3rd quartile class

f_3 = frequency of the 3rd quartile class

m_3 = c.f. preceding the 3rd quartile class

c_3 = width of the third quartile class

Example 5.32

Calculate the quartiles Q_1 and Q_3 for wages of the labours given below

Wages (Rs.)	30-32	32-34	34-36	36-38	38-40	40-42	42-44
Labourers	12	18	16	14	12	8	6

Solution:

x	f	cf
30 – 32	12	12
32 – 34	18	30
34 – 36	16	46
36 – 38	14	60
38 – 40	12	72
40 – 42	8	80
42 – 44	6	86
	86	

$$\frac{N}{4} = \frac{86}{4} = 21.5$$

lies in the group 32 – 34

$$Q_1 = l_1 + \frac{\frac{N}{4} - m_1}{f_1} \times c_1$$

$$= 32 + \frac{21.5 - 12}{18} \times 2$$

$$= 32 + \frac{19}{18} = 32 + 1.06$$

$$= \text{Rs. } 33.06$$

$$\frac{3N}{4} = \frac{3 \times 86}{4} = 64.5$$

$\therefore Q_3$ lies in the group 38 – 40

$$Q_3 = l_3 + \frac{3\left(\frac{N}{4}\right) - m_3}{f_3} \times C_3$$

$$= 38 + \frac{64.5 - 60}{12} \times 2$$

$$= 38 + \frac{4.5}{12} \times 2$$

$$= 38 + 0.75 = \text{Rs. } 38.75$$

5.5.2 Deciles

Deciles are similar to quartiles. Quartiles divides ungrouped data into four quarters and Deciles divide data into 10 equal parts .

Example 5.33

Find the D_6 for the following data

11, 25, 20, 15, 24, 28, 19, 21

Solution:

Arrange in an ascending order

11,15,19,20,21,24,25,28

$$D_6 = \left(\frac{6(n+1)}{10}\right)^{th} \text{ item}$$

$$= \left(\frac{6(8+1)}{10}\right)^{th} \text{ item}$$

$$= \left(\frac{6(9)}{10}\right)^{th} \text{ item}$$

$$= [5.4]^{th} \text{ item}$$

$$= 5^{th} \text{ item} + (0.4)\,(6^{th} \text{ item} - 5^{th} \text{ item})$$

$$D_6 = 21+(0.4)(24-21)$$

$$= 21+(0.4)(3)$$

$$= 21+1.2$$

$$= 22.2$$

Example 5.34

Calculate D_5 for the frequency distribution of monthly income of workers in a factory

Income (in thousands)	0 – 4	4 – 8	8 – 12	12 – 16	16 – 20	20 – 24	24 – 28	28 – 32
No of persons	10	12	8	7	5	8	4	6

Solution:

Class	f	cf
0-4	10	10
4-8	12	22
8-12	8	30
12-16	7	37
16-20	5	42
20-24	8	50
24-28	4	54
28-32	6	60
	N=60	

$$D_5 = \left(\frac{5N}{10}\right)^{th} \text{ item}$$

$$= \left(\frac{5(60)}{10}\right)^{th} \text{ item}$$

$$= 30^{th} \text{ item}$$

This item in the interval 8–12

$l = 8, \ m = 22, \ f = 8, \ c = 4, \ N = 60$

$$D_5 = l + \left(\frac{\frac{5N}{10} - m}{f}\right) \times c$$

$$= 8 + \left(\frac{30 - 22}{8}\right) \times 4$$

$$= 8 + \frac{8}{8} \times 4$$

$$D_5 = 12$$

5.5.3 Percentiles

The percentile values divide the frequency distribution into 100 parts each containing 1 percent of the cases. It is clear from the definition of quartiles, deciles and percentiles

Relationship

$$P_{25} = Q_1$$

$$P_{50} = \text{Median} = Q_2$$

$$P_{75} = 3^{rd} \text{ quartile} = Q_3$$

Example 5.35

The following is the monthly income (in 1000) of 8 persons working in a factory. Find P_{30} income value

10,14, 36, 25, 15, 21, 29, 17

Solutions:

Arrange the data in an ascending order.

$$n = 8$$

10,14,15,17,21,25,29,36

$$P_{30} = \left(\frac{30(n+1)}{100}\right)^{th} \text{ item}$$

$$= \left(\frac{30 \times 9}{100}\right)^{th} \text{ item}$$

$$= 2.7^{th} \text{ item}$$

$$= 2^{nd} \text{ item} + 0.7\ (3^{rd} \text{ items} - 2^{nd} \text{ items})$$

$$= 14+0.7(15\text{-}14)$$

$$= 14+0.7$$

$$P_{30} = 14.7$$

Example 5.36

Calculate P_{61} for the following data relating to the height of the plants in a garden

Heights (in cm)	0 – 5	5 – 10	10 – 15	15 – 20	20 – 25	25 – 30
No of plants	18	20	36	40	26	16

Solution:

Class	*f*	*cf*
0 – 5	18	18
5 – 10	20	38
10 – 15	36	74
15 – 20	40	114
20 – 25	26	140
25 – 30	16	156
	N=156	

$$\frac{61N}{100} = \left(\frac{61 \times 156}{100}\right)$$

$$= 95.16 \text{ item}$$

This item is the interval 15-20. Thus

$l = 15$, $m = 74$, $f = 40$, $c = 5$

$$P_{61} = l + \left(\frac{\frac{61N}{100} - m}{f}\right) \times c$$

$$= 15 + \left(\frac{95.16 - 74}{40}\right) \times 5$$

$$= 15 + \frac{21.16}{40} \times 5$$

$$= 17.645$$

$$P_{61} = 17.645$$

Points to Remember

- A central tendency is a single figure that represents the whole mass of data
- Arithmetic mean or mean is the number which is obtained by adding the values of all the items of a series and dividing the total by the number of items.
- When all items of a series are given equal importance than it is called simple arithmetical mean and when different items of a series are given different weights according with their relative importance is known as weighted arithmetic mean.

- Median is the middle value of the series when arranged in ascending order
- When a series is divided into more than two parts, the dividing values are called partition values.
- Mode is the value which occurs most frequently in the series,that is modal value has the highest frequency in the series.

EXERCISE 5

I. Choose the best answer:

1. Which of the following is a measure of central value?

   (a) Median (b) Deciles

   (c) Quartiles (d) Percentiles

2. Geometric Mean is better than other means when

   (a) the data are positive as well as negative

   (b) the data are in ratios or percentages

   (c) the data are binary

   (d) the data are on interval scale

3. When all the observations are same, then the relation between A.M., G.M. and H.M. is:

   (a) A.M. = G.M. = H.M. (b) A.M. $<$ H.M. $<$ G.M.

   (c) A.M. $<$ G.M. $<$ H.M. (d) A.M. $>$ G.M. $>$ H.M.

4. The median of the variate values 11, 7, 6, 9, 12, 15, 19 is

   (a)9 (b)12 (c)15 (d)11

5. The middle values of an ordered series is called

   (a) 50^{th} percentile (b) 2^{nd} quartile

   (c) 5^{th} decile (d) all the above

6. Mode is that value in a frequency distribution which possesses

   (a) minimum frequency (b) maximum frequency

   (c) frequency one (d) none of the above

7. For decile, the total number of partition values are:

(a) 5 (b) 8 (c) 9 (d) 10

8. The mean of the squares of first eleven natural numbers is:

(a) 46 (b) 23 (c) 48 (d) 42

9. Histogram is useful to determine graphically the value of

(a) mean (b) median (c) mode (d) all the above

10. What percentage of values lies between 5^{th} and 25^{th} percentile?

(a) 15% (b) 30% (c) 75% (d) none of the above

II. Fill in the blanks

11. In an open end distribution ____________ cannot be determined.

12. The sum of the deviations from mean is _____________

13. The distribution having two modes is called ___________

14. Second quartile and ___________ deciles are equal.

15. Median is a more suited average for grouped data with __________ classes.

III. Very Short Answer Questions:

16. What is meant by measure of central tendency?

17. What are the desirable characteristics of a good measure of central tendency?

18. What are the merits and demerits of the arithmetic mean?

19. Express weighted arithmetic mean in brief.

20. Define Median. Discuss its advantages and disadvantages.

IV. Short Answer Questions :

21. The monthly income of ten families of a certain locality is given in rupees as below.

Family	A	B	C	D	E	F	G	H	I	J
Income (₹)	85	70	10	75	500	8	42	250	40	36

Calculate the arithmetic mean.

22. The mean of 100 items are found to be 30. If at the time of calculation two items are wrongly taken as 32 and 12 instead of 23 and 11. Find the correct mean.

23. A cyclist covers his first three kms at an average speed of 8 kmph. Another two kms at 3 kmph and the last two kms at 2 kmph. Find the average speed for the entire journey.

24. The mean marks of 100 students were found to 40. Later it was discovered that a score of 53 was misread as 83. Find the corrected mean corresponding to the corrected score.

25. In a moderately asymmetrical distribution the values of mode and mean are 32.1 and 35.4 respectively. Find the median value.

26. Calculate D_9 from the following frequency distribution

x	58	59	60	61	62	63	64	65	66
f	2	3	6	15	10	5	4	3	2

27. Calculated P_{40} The following is the distribution of weights of patients in an hospital

Weight (in kg)	40	50	60	70	80	90	100
No of patients	15	26	12	10	8	9	5

V. Calculate the following:

28. Find the mean and median:

Wages (₹)	60 – 70	50 – 60	40 – 50	30 – 40	20 – 30
No. of labourers	5	10	20	5	3

29. The following data relates to the marks obtain by students in a school find the median

Marks	>10	>20	>30	>40	>50	>60	>70	>80	>90
No. of Students	70	62	50	38	30	24	17	9	4

30. The number of telephone calls received in 245 successive one minute intervals at an exchange are shown in the following frequency distribution:

No. of calls	0	1	2	3	4	5	6	7
Frequency	14	21	25	43	51	40	39	12

Evaluate the mean, median and mode.

31. Calculate the geometric and the harmonic mean of the following series of monthly expenditure of a batch of students.

125, 130, 75, 10, 45, 0.5, 0.40, 500, 150, 5

32. Find out the mode of the following series:

Wages (₹)	Below 25	25 – 50	50 – 75	75 – 100	100 – 125	Above 125
No. of persons	10	30	40	25	20	15

33. Find the median, lower quartile, 7th decile and 85th percentile of the frequency distribution given below:

Marks in Statistics	Below 10	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60	60 – 70	Above 70
No. of Students	8	12	20	32	30	28	12	4

34. The following data gives the distribution of heights of a group of 60 college students:

Height (in cms)	Number of students
145.0 – 149. 9	2
150.0 – 154.9	5
155.0 – 159.9	9
160.0 – 164.9	15
165.0 – 169.9	16
170.0 – 174.9	7
175.0 – 179.9	5
180.0 – 184.9	1

Draw the histogram for this distribution and find the modal height. Check this result by using the algebraic formula.

35. Find the median and the quartiles for the following datas:

Size	Frequency	Size	Frequency
4	40	12	50
5	48	13	52
6	52	14	41
7	56	15	57
8	60	16	63
9	63	17	52
10	57	18	48
11	55	19	43

36. In a class of 50 students, 10 have failed and their average of marks is 2.5. The total marks secured by the entire class were 281. Find the average marks of those who have passed.

Answers

I. 1. (a) 2. (b) 3. (a) 4. (d) 5. (d) 6. (b) 7. (c) 8. (a) 9. (c) 10. (d)

II. 11. Mean 12. Zero 13. Bimodal 14. 5th

15. Open end IV. 21. Mean = 111.60 22. Mean = 29.90

23. average speed = 3.4 mph 24. corrected mean = 39.7

25. Median = 34.3 26. $D_9 = 65$ 27. $P_{40} = 50$

V. 28. Mean = 47.09, Median = 46.75 29. Median = 43.75

30. Mean = 3.76, Median = 4, Mode = 4

31. G.M. = 57.73 and H.M. = 2.058 32. Mode = Rs. 60

33. Median = 40.33, First quartile = 28.25, 7th decile = 50.07, 85th percentile = 57.9

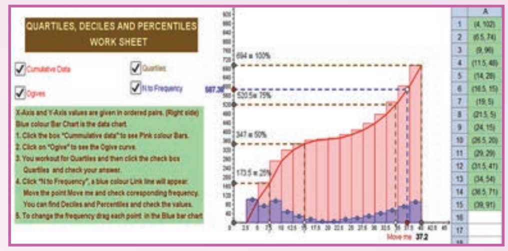
35. Median = 11, First quartile = 8, Third quartile = 15

36. Average marks of those who passed = 6.4

ICT CORNER

QUARTILES, DECILES, PERCENTILES

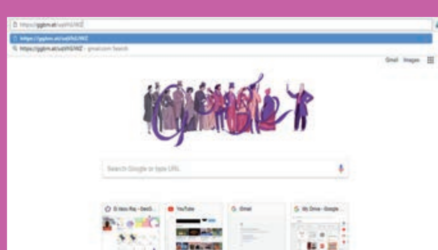
This activity helps to understand about quartiles, deciles and percentiles diagrammatically



Steps:

- Open the browser and type the URL given (or) scan the QR code.
- GeoGebra work book called "**11th Standard Statistics**" will appear. In this several work sheets for Statistics are given, Open the worksheet named "**Quartiles**"
- Quartiles page will open. In that blue colour histogram is seen. If you click on "**Cumulative data**" pink colour chart for "**Cumulative frequency**" will appear. If you click on "**Ogive**" red colour ogive curve will appear. Now you calculate the quartiles, deciles and percentiles using the formula given in the text book.
- Now If press "**Quartiles**" check box to see the quartiles and check your answer. Now press "**N to frequency**" check box and move "**Blue dot**" on the x-axis to check your deciles and percentiles. If necessary You can create your own problem by typing new frequencies

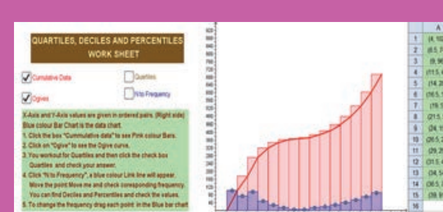
Step-1



Step-2

Step-3

Step-4



Pictures are indicatives only*

URL:

https://ggbm.at/uqVhSJWZ