Chapter 5 (Notes)

5 CENTRAL TENDENCY

Arthur Lyon Bowley

Please share and subscribe ”stati4u”

Desirable properties of a good average

An average should posses the following properties.

(i) Simple and rigid definition.

(ii) Simple to understand and easy to calculate.

(iii) Based on all the observations.

(iv) Least affected by extreme values.

(v) Least affected by fluctuations of sampling.

(vi) Capable of further mathematical treatment.

Various measures of Central Tendencies

1. Arithmetic Mean (AM)
2. Median
3. Mode
4. Geometric Mean (GM)
5. Harmonic Mean (HM)

Arithmetic Mean (AM)

Computation of Arithmetic Mean

(i) Arithmetic Mean from a raw data

$$\bar{x} = \frac{\text{sum of the observations}}{\text{number of observations}}$$

$$= \frac{x_1 + x_2 + x_3 + \cdots + x_n}{n}$$

$$= \frac{\sum x}{n}$$

$$\text{Mean} = \frac{\sum x}{n}$$

The Greek letter $\sum$ represents summation

Remark: The sum of the observations in a series, $\sum x = n\bar{x}$

Illustration 5.1

An umbrella manufacturing company wants to launch a new product in a state. The rainfall (in cms) in the state for the last five years is, 120, 135, 110, 142 and 150. Find the average rainfall of the state for the last five years.

Solution. Average rainfall,

$$\bar{x} = \frac{\sum x}{n}$$

$$= \frac{120 + 135 + 110 + 142 + 150}{5}$$

$$= 131.4\,cm$$

Illustration 5.2

In the first four class tests, a student got the scores 52, 48, 33 and 27 respectively. (a) Find the mean of the scores.
(b) If in the fifth test, he got a score of 45, find his new mean score.

Solution. (a) The mean score of the first four class tests is

$$\bar{x} = \frac{\sum x}{n}$$

$$= \frac{52 + 48 + 33 + 27}{4}$$
$$= 40$$

(b) The new mean after 5 class tests is

$$\bar{x} = \frac{\sum x}{n}$$

$$= \frac{52 + 48 + 33 + 27 + 45}{5}$$

$$= 41$$

Illustration 5.3

The mean of a group of 100 observations is known to be 50. Later it was discovered that two observations were misread as 92 and 8 instead of 192 and 88. Find the correct mean.

Solution. Given that $\bar{x} = 50$ and $n = 100$.

We have the sum of the observations,

$$\begin{aligned} \sum x &= n\bar{x} \\ &= 100 \times 50 \\ &= 5000 \end{aligned}$$

But it was wrong, because two observations 192 and 88 were misread as 92 and 8. So the corrected sum of observations is

$$\begin{aligned} \text{corrected sum} &= 5000 - 92 - 8 + 192 + 88 \\ &= 5180 \end{aligned}$$

$$\begin{aligned} \text{So corrected mean,} \bar{x} &= \frac{\text{corrected sum}}{n} \\ &= \frac{5180}{100} \\ &= 51.80 \end{aligned}$$

Arithmetic Mean from a Discrete Frequency Distribution

$$\bar{x} = \frac{\sum f x}{\sum f}$$

ie.$\bar{x} = \frac{\sum f x}{N}$ where $N = \sum f$ is the total frequency

Illustration 5.4

The students in a Statistics class were trying to study the heights of participants in a sports meet. They collected the height of 20 participants, as displayed in the table.

Height (in inches)	:	49	53	54	55	66	68	70	80
No of participants	:	1	2	4	5	3	2	2	1

Calculate the mean height of the participants.

Solution. Average height, $\bar{x} = \dfrac{\sum f x}{N}$ where $N = \sum f$

Height (x)	No of participants (f)	$f x$
49	1	49
53	2	106
54	4	216
55	5	275
66	3	198
70	2	140
80	1	80
Total	$N = 20$	1200

$$\bar{x} = \frac{\sum f x}{N}$$
$$= \frac{1200}{20}$$
$$= 60$$

Illustration 5.5

A survey is taken by an insurance company to determine how many car accidents the average New Delhi City resident has gotten into in the past 10 years. The company surveyed 200 people who are getting off a train at a subway station. The following table gives the results of the survey.

Number of Accidents	Number of People
0	60
1	10
2	40
3	10
4	80

Calculate the mean number of accidents of this data set.

Solution. Average height, $\bar{x} = \frac{\sum fx}{N}$ where $N = \sum f$

No of accidents (x)	No of people (f)	fx
0	70	0
1	10	10
2	40	80
3	10	30
4	70	280
Total	$N = 200$	400

$$\bar{x} = \frac{\sum fx}{\sum f} = \frac{400}{200} = 2$$

Arithmetic Mean from a continuous frequency distribution

A continuous frequency distribution consists of data that are grouped by classes.

$$\bar{x} = \frac{\sum fx}{N} \text{ where } N = \sum f$$

The table below show the age of 55 patients selected to study the effectiveness of a particular medicine.

Age	No of Patients
0-10	5
10-20	7
20-30	17
30-40	12
40-50	5
50-60	2
60-70	7

Calculate the mean age of the patients.

Solution. Mean age,

$$\bar{x} = \frac{\sum f x}{N} \text{ where } N = \sum f$$

To find mean, prepare the following table

Age	Mid point (x)	No of patients (f)	fx
0-10	5	5	25
10-20	15	7	105
20-30	25	17	425
30-40	35	12	420
40-50	45	5	225
50-60	55	2	110
60-70	65	7	455
Total		$N = 55$	1765

$$\bar{x} = \frac{\sum f x}{\sum f} = \frac{1765}{55} = 32.09$$

The frequency distribution below represents the weights in kg of parcels carried by a small logistic company. Find the mean weight of parcels.

Weight	No. of Parcels
10.0-10.9	2
11.0-11.9	3
12.0-12.9	5
13.0-13.9	8
14.0-14.9	12
15.0-15.9	15
16.0-16.9	13
17.0-17.9	11
18.0-18.9	6
19.0-19.9	2

Solution. Mean weight,

$$\bar{x} = \frac{\sum f x}{N} \text{ where } N = \sum f$$

Weight	Mid point (x)	No. of Parcels (f)	fx
10.0-10.9	10.45	2	20.90
11.0-11.9	11.45	3	34.35
12.0-12.9	12.45	5	62.25
13.0-13.9	13.45	8	107.60
14.0-14.9	14.45	12	173.40
15.0-15.9	15.45	15	231.75
16.0-16.9	16.45	13	213.85
17.0-17.9	17.45	11	191.95
18.0-18.9	18.45	6	110.70
19.0-19.9	19.45	2	38.90
Total		$N = 67$	1185.65

$$\bar{x} = \frac{\sum fx}{\sum f}$$
$$= \frac{1185.65}{67}$$
$$= 17.70$$

Mean weight = 17.70 kg

Mathematical Properties of Arithmetic Mean

The AM of a distribution has the following mathematical properties.

1. The sum of deviations of items in a data from the AM is always Zero.

$$ie \sum (x - \bar{x}) = 0$$

2. The sum of squares of the deviations of the items in a data is the least when the deviation is taken about the Mean.

$$ie \sum (x - a)^2 \text{ is least when } a = \bar{x}$$

3. If the mean of n observations, $x_1, x_2, \ldots, x_n$ is $\bar{x}$ then the mean of the observations, $(x_1 \pm a), (x_2 \pm a), \ldots, (x_n \pm a)$ is $(\bar{x} \pm a)$.

ie, If each observation is increased by 'a', then the mean also increased by a and if each observation is decreased by 'a', then the mean is also decreased by a.

4. The mean of n observations, $x_1, x_2, \ldots, x_n$ is $\bar{x}$. If each observation is multiplied by $p, p \neq 0$, then the mean of the new observations is $p\bar{x}$.

Merits and Demerits of AM

Merits

1. It has a rigid definition.
2. It is easy to calculate and understand.
3. AM is based upon all the observations.
4. It is least affected by fluctuations of sampling.
5. It is capable of further mathematical treatment.

Demerits

1. AM is highly affected by extreme values.

2. It can’t be determined by inspection.

3. It can’t be used for qualitative characteristics like, intelligence, honesty, beauty etc.

4. It can’t be calculated for open end classes.

Weighted Arithmetic Mean

Computation of Weighted Arithmetic Mean

Let $x_1, x_2, \ldots, x_n$ be the observations of a data and $w_1, w_2, \ldots, w_n$ be their corresponding weights. Then the weighted arithmetic mean is given by,

$$\bar{x} = \frac{w_1 x_1 + w_2 x_2 + \cdots + w_n x_n}{w_1 + w_2 + \cdots + w_n} = \frac{\sum wx}{\sum w}$$

Illustration 5.8

A student's final scores in Mathematics, Physics, Chemistry and English are respectively 82, 86, 90 and 70. If the respective credits received for these courses are 3, 5, 3, and 1, determine the average score.

Solution. Here the weights associated to the observations 82. 86. 90 and 70 are 3, 5, 3 and 1.

x	:	82	86	90	70
w	:	3	5	3	1

Average,

$$\bar{x} = \frac{\sum wx}{\sum w}$$

$$= \frac{82 \times 3 + 86 \times 5 + 90 \times 3 + 70 \times 1}{3 + 5 + 3 + 1}$$

$$= \frac{1016}{12}$$

$$= 84.67$$

Combined Arithmetic Mean

If $\bar{x_1}$ and $\bar{x_2}$ are the means of two groups of n_1 and n_2 observations respectively, the mean of the combined group of $n_1 + n_2$ observations is given by

$$\bar{x} = \frac{n_1\bar{x_1} + n_2\bar{x_2}}{n_1 + n_2}$$

Illustration 5.9

The mean score obtained in an examination by a group of 100 students was found to be 50. The mean of the scores obtained in the same examination by another group of 200 students was 57. Find the mean of scores obtained by both the groups taken together.

Solution. We are given that

$$\bar{x}_1 = 50 \text{ and } \bar{x}_2 = 57$$

$$n_1 = 100 \text{ and } n_2 = 200$$

We know that the combined mean is given by

$$\bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2}$$

$$= \frac{100 \times 50 + 200 \times 57}{100 + 200}$$

$$= 54.67$$

Illustration 5.10

The mean weight of 150 students in a certain class is 60 kgs. The mean weight of boys in the class is 70 kgs and that of girls is 55 kgs. Find the number of boys and number of girls in the class.

Solution. We are given that

The combined mean, $\bar{x} = 60\,\text{kgs}$

Mean weight of boys, $\bar{x_1} = 70\,\text{kgs}$

Mean weight of girls, $\bar{x_2} = 55\,\text{kgs}$

The total no of students $= 150$

Let there be 'x' boys in the class. Therefore the number of girls in the class is $150 - x$. We know that,

$$\bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2}$$

$$\text{ie,}60 = \frac{70x + 55(150 - x)}{150}$$

$$\Rightarrow 9000 = 70x + 8250x - 55x$$

$$\Rightarrow 15x = 750 \Rightarrow x = 50$$

So number of boys in the class is 50 and number of girls in the class is $150 - 50 = 100$.

If there are k groups of sizes $n_1, n_2, \ldots\ldots, n_k$ respectively and $\bar{x_1}, \bar{x_2}, \ldots\ldots, \bar{x_k}$ are their respective means, then the combined mean of $n_1 + n_2 + \ldots + n_k$ observations is given by

$$\bar{x} = \frac{n_1\bar{x_1} + n_2\bar{x_2} + \ldots + n_k\bar{x_k}}{n_1 + n_2 + \ldots + n_k}$$

Median

Median is the value of the middlemost observations in the data when the data is arranged in ascending or descending order.

(i) Calculation of Median from a raw data

Consider a raw data having 'n' observations. To find the median, first arrange the data in ascending or descending order. Then median is the $\left(\frac{n+1}{2}\right)^{\text{th}}$ item in the data.

> Median of a raw data is the $\left(\frac{n+1}{2}\right)^{\text{th}}$ item, when the data is arranged in ascending or descending order of magnitude.

Illustration 5.11

Rahna's maths quiz scores in 9 competitions were 88, 97, 87, 92, 90, 88, 93, 98 and 95. What was her median quiz score ?

Solution. Arranging the data in ascending order,

$$87, 88, 88, 90, 92, 93, 95, 96, 98$$

Here the number of observations, $n = 9$.

$$\text{Median} = \left(\frac{n+1}{2}\right)^{\text{th}} \text{ item}$$

$$= \left(\frac{9+1}{2}\right)^{\text{th}} \text{ item}$$

$$= 5^{\text{th}} \text{ item}$$

The 5^{th} item in the series is 92. Median quiz score is 92

Illustration 5.12

Anand's family plans a trip from Thiruvananthapuram to Wayanad on their summer vacation. They drove through 8 districts. The following are the petrol prices in the 8 districts on those days. Rs.71.9, Rs.72.3, Rs.72.4, Rs. 72.32, Rs. 73, Rs.73.1, Rs.72.2 and Rs.72.48 What is the median petrol price ?

Solution. Arranging the observations in ascending order,

$$71.9,\ 72.2,\ 72.3,\ 72.32,\ 72.4,\ 72.48,\ 73,\ 73.1$$

Number of observations is 8.

$$\text{Median} = \left(\frac{n+1}{2}\right)^{\text{th}} \text{ item}$$

$$= \left(\frac{8+1}{2}\right)^{\text{th}} \text{ item}$$

$$= 4.5^{\text{th}} \text{ item}$$

But there is no item in the series having a position 4.5. So we take the mean of the 4^{th} and 5^{th} items in the series as the median.

$$\text{Median} = \text{Mean of } 4^{\text{th}} \text{ and } 5^{\text{th}} \text{ item}$$

$$= \frac{72.32 + 72.4}{2}$$

$$= 72.36$$

Median petrol price = Rs. 72.36.

(ii) Median from a Discrete Frequency Distribution

For a discrete frequency distribution median is the observation having cumulative frequency $\frac{N+1}{2}$, when the observations are arranged in ascending order

The following steps can be used to find the median in a discrete distribution.

Step 1: Arrange the data in ascending or descending order of magnitude.

Step 2: Obtain the cumulative frequencies.

Step 3: Determine $\frac{N+1}{2}$, where N is the total frequency.

Step 4: Median is the value for the $\left(\frac{N+1}{2}\right)^{th}$ item of the data.

Illustration 5.13

The following data gives the daily wages of workers in a manufacturing company. Find the median wage.

Daily wage (in 100 Rupees)	:	6	8	10	12	15	18
Number of workers		20	14	7	16	12	2

Solution. The given data is in ascending order.

The cumulative frequencies are given by,

Wages (in 100 rupees)	Number of workers (frequency)	Cumulative frequency
6	20	20
8	14	34
10	7	41
12	16	57
15	12	69
18	2	71
	$N = 71$	

Total frequency $N = 71$. So $\frac{N+1}{2} = \frac{72}{2} = 36$

$\therefore$ Median is the value in the data which comes in the 36^{th} position. Which is the value of the item having cumulative frequency 36, which is 10.

$\therefore$ Median = Rs. 1000/-

Illustration 5.14

The table below shows the marks obtained by 42 students in an examination.

Marks	9	20	25	40	50	80
Number of students	4	6	11	13	7	2

Calculate the median mark.

Solution. The data given is in ascending order. The cumulative frequencies are

Marks	Frequency	Cumulative frequency
9	4	4
20	6	10
25	11	21
40	12	33
50	7	40
80	2	42
	$N = 42$	

Total frequency, $N = 42$.

$$\frac{N+1}{2} = \frac{43}{2} = 21.5$$

$\therefore$ Median is the value in the data which comes in the 21.5^{th} position. But there is no observation in the data which has the cumulative frequency, 21.5. Hence we consider the median as the mean of the 21^{th} and 22^{nd} observations. The 21^{th} observation is 25 and 22^{nd} observation is 40

$$\text{ie, Median} = \frac{25+40}{2} = \frac{65}{2} = 32.5$$

(iii) Median from a Continuous Frequency Distribution

To find median in continuous frequency series, first we have to locate median class. Median class is the class where $(\frac{N}{2})^{th}$ observation lies. Median of a continuous frequency series is given by

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

Where l - lower bound (actual class limit) of the median class.

c - class interval of the median class.

f - frequency of the median class and

m - cumulative frequency of the class preceding the median class.

The following steps can be used to determine the median for a continuous frequency series.

Step 1: Convert the inclusive type classes to the exclusive type classes (if any).

Step 2: Obtain the cumulative frequencies.

Step 3: Determine $\frac{N}{2}$, where N is the total frequency.

Step 4: Locate the class having cumulative frequency $\frac{N}{2}$.

Step 5: Find median using the above formula.

(iii) Median from a Continuous Frequency Distribution

To find median in continuous frequency series, first we have to locate median class. Median class is the class where $(\frac{N}{2})^{th}$ observation lies. Median of a continuous frequency series is given by

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

Where l - lower bound (actual class limit) of the median class.

c - class interval of the median class.

f - frequency of the median class and

m - cumulative frequency of the class preceding the median class.

The following steps can be used to determine the median for a continuous frequency series.

Step 1: Convert the inclusive type classes to the exclusive type classes (if any).

Step 2: Obtain the cumulative frequencies.

Step 3: Determine $\frac{N}{2}$, where N is the total frequency.

Step 4: Locate the class having cumulative frequency $\frac{N}{2}$.

Step 5: Find median using the above formula.

The distribution of income of 63 families is,

Income (100 Rupees)	:	30-40	40-50	50-60	60-70	70 -80	80-90	90 -100
No of workers	:	6	12	18	13	9	4	1

Compute the median income.

Solution. The cumulative frequency table is

Class	Frequency	Cumulative frequency
30-40	6	6
40-50	12	18
50-60	18	36
60-70	13	49
70-80	9	58
80-90	4	62
90-100	1	63
	$N = 63$	

$$\frac{N}{2} = 31.5$$

The class having cumulative frequency 31.5 is 50-60.

∴ Median class is 50-60.

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

Where, $l = 50, c = 10, f = 18$ and $m = 18$.

$$Median = 50 + \frac{(31.5 - 18)10}{18}$$

$$= 50 + \frac{135}{18}$$

$$= 57.5$$

Illustration 5.16

The table below shows the distribution of marks obtained by 50 students in Economics. Find the median mark.

Marks	:	*10-14*	*15-19*	*20-24*	*25-29*	*30-34*	*35-39*
Number of students	:	*4*	*6*	*10*	*5*	*7*	*3*
	:	*40-44*	*45-49*				
	:	*9*	*6*				

Solution. Here the classes are of inclusive type. Before computing median, we have to convert it into exclusive form to get the actual class limits (class bounds).

Let us prepare the cumulative frequency table with the actual class limits.

Marks	Actual class	Frequency	Cumulative frequency
10-14	9.5-14.5	4	4
15-19	14.5-19.5	6	10
20-24	19.5-24.5	10	20
25-29	24.5-29.5	5	25
30-34	29.5-34.5	7	32
35-39	34.5-39.5	3	35
40-44	39.5-44.5	9	44
45-49	44.5-49.5	6	50
		$N = 50$	

Here $N = 50. \therefore \frac{N}{2} = 25$

Hence the median class is $24.5 - 29.5$.

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

Where $l = 24.5, c = 5, f = 5$ and $m = 20$

$$Median = 24.5 + \frac{(25 - 20)5}{5}$$

$$= 29.5$$

Illustration 5.17

The table below shows the marks obtained by 100 students in an examination. Locate median graphically.

Marks	:	*0-10*	*10-20*	*20-30*	*30-40*	*40-50*	*50-60*	*60-70*
No of students	:	*7*	*10*	*21*	*27*	*22*	*9*	*4*

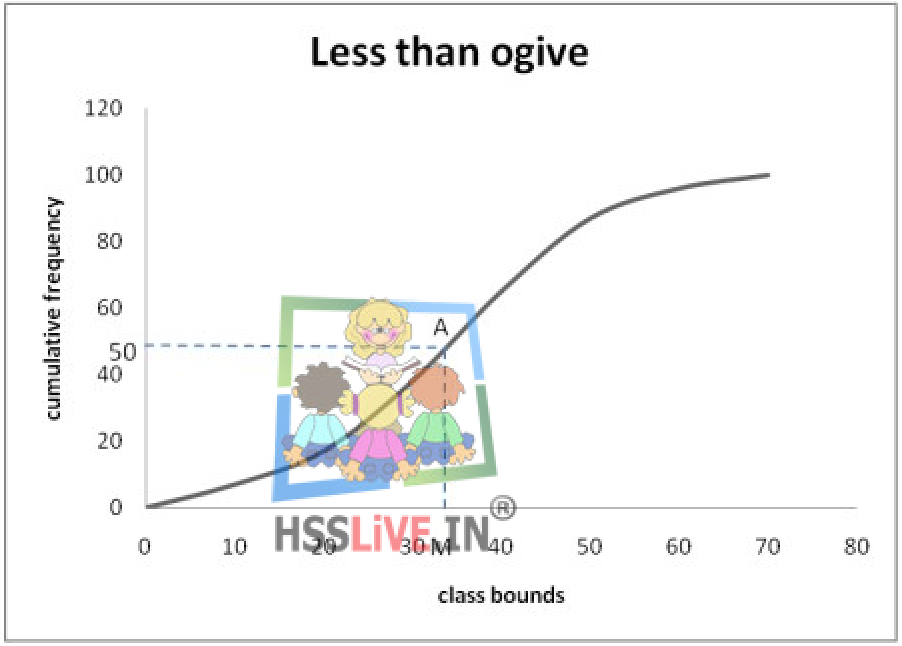
Solution.

(i) Median using 'less than' ogive.

In order to find median using 'less than' ogive, we have to construct the less than frequency table.

Upper bound	Less than cumulative frequency
10	7
20	17
30	38
40	65
50	87
60	96
70	100

$N = 100$ so that $\frac{N}{2} = 50$. Draw a less than ogive.


Less than ogive
cumulative frequency
120
100
80
60
50
40
20
0
A
0
10
20
30 M
40
50
60
70
80
class bounds

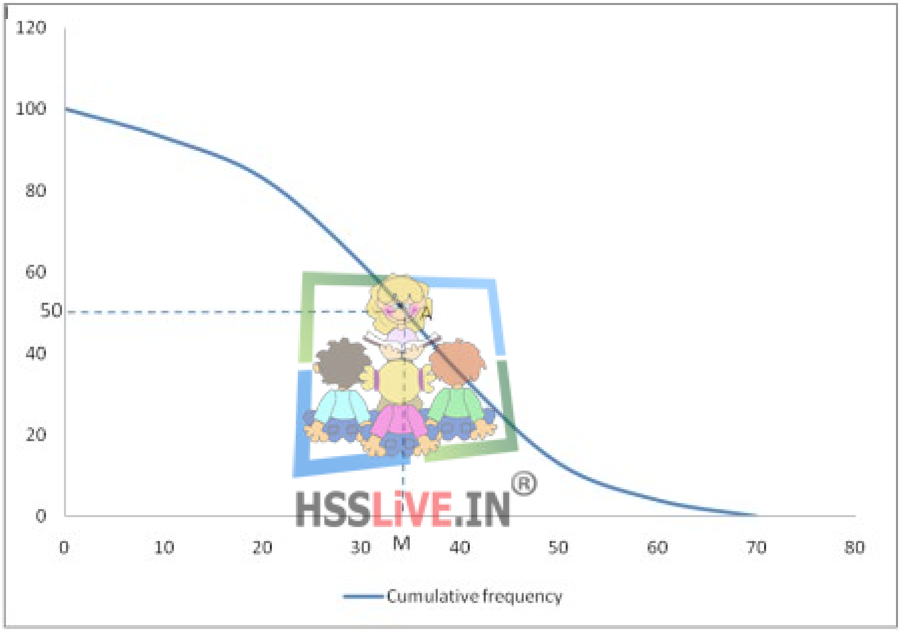
∴ Median is 34.

(ii) Median using more than ogive.

We have to first prepare a greater than cumulative frequency table.

Lower bound	Greater than cumulative frequency
0	100
10	93
20	83
30	62
40	35
50	13
60	4

Draw a more than ogive.



∴ Median is 34.

(ii) Median from two ogives ('less than' and 'more than' ogives).

The following are the steps involved in the determination of median from 'less than' and 'more than' ogives by simultaneously drawing them.

Step 1: Draw the two ogives on a paper with the same axis.

Step 2: Mark the point A, where the two ogives intersect.

Step 3: Draw perpendicular from A to the x-axis. The corresponding value on the x-axis would be the median of the data.

Illustration 5.18

Determine the Median using the data given in ***Illustration 5.17***

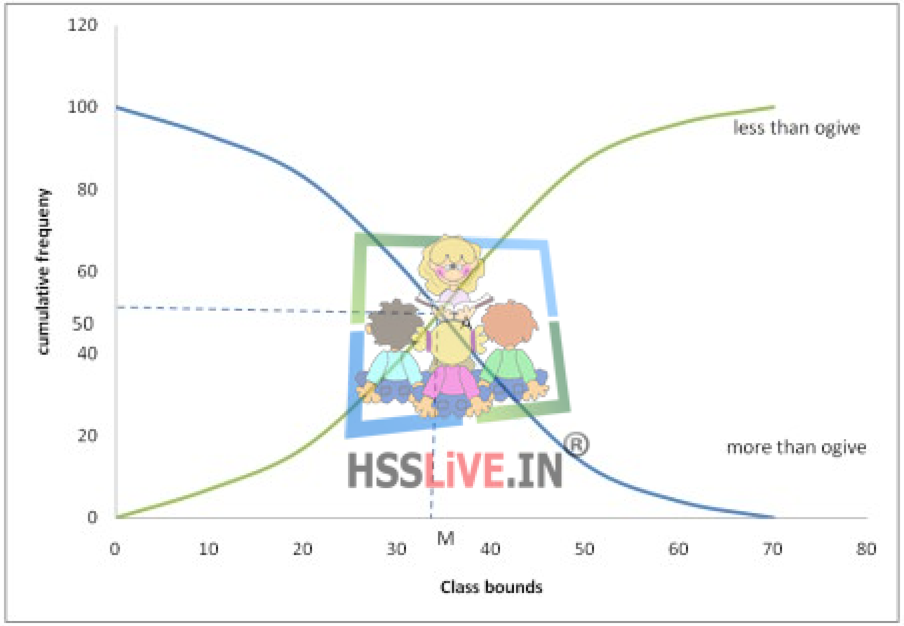
Marks	:	0-10	10-20	20-30	30-40	40-50	50-60	60-70
No of students	:	7	10	21	27	22	9	4

Solution. Prepare the two cumulative frequency tables.

Upper bound	Less than cumulative frequency	Lower bound	Greater than cumulative frequency
10	7	0	100
20	17	10	93
30	38	20	83
40	65	30	62
50	87	40	35
60	96	50	13
70	100	60	4

Draw the two ogives simultaneously on the same paper as shown below.

Draw the two ogives simultaneously on the same paper as shown below.



Median = 34.

5.2.2 Merits and demerits of Median

Merits

1. It has a rigid definition.
2. Median is easy to compute. In some cases it be located merely by inspection.
3. It is not affected by the extreme values.

4. It can be calculated for distribution with open end classes.

5. It is the only measure to be used while dealing with qualitative data which can measure quantitatively but can be arranged ascending or descending order of magnitudes.

6. The median may be a better indicator if a set of numbers has an *outlier*. An *outlier* is an extreme value that differs greatly from the other values.

Demerits

1. In some cases median can't be calculated exactly. For example, in the case of even number of observations, we take median as the mean of the two middle terms, as an approximation.
2. It is not based on all the observations.
3. It is affected by the fluctuations of sampling.
4. It is not capable of further mathematical treatments.

Mode

Mode of a data is defined as the value that is repeated most often in the data. It is the observation having the maximum frequency in a data.

Mode is sometimes called the *Fashionable Average* or *Business Average*

(i) Calculating the mode from a raw data.

For a raw data mode is the value which appears most often in the data. ie Mode of a raw data is the observation which appears a maximum number of times in the data.

Illustration 5.19

The ages of six persons who participated in an interview were 20, 21, 21, 24, 25, 24, 21 *and 27 years. Find the mode of the data.*

Solution. Here the observation 21 appears three times, 24 appears two and all others are appear in a single time. So the value which appears a maximum number of times is 21.

$\therefore$ Mode = 21 years.

In a distribution, there may be one, two or more than two modes. If in a distribution there are two observations which are appearing a maximum number of times, then both the observations are taken as the modes of the distribution. A distribution which has a single mode is called a unimodal distribution, which has two modes is called a bimodal and has more than two modes is called a multimodal distribution.

Illustration 5.20

Mr. Vijayakumar, the physical education teacher of a school is trying to determine the average height of students in the cricket team of the school. The height of the players in inches are 70, 72, 72, 74, 74, 74, 75, 76, 76, 76 and 77. Calculate the mode of the heights.

Solution.
Here the data has two values, 74 and 76, which appears 3 times.
All the other values appear less than 3 times.So the data set has two distinct modes 74 and 76.

(ii) Mode from a discrete frequency distribution

Mode of a discrete frequency distribution is the observation having the maximum frequency.

> The mode of a discrete frequency is the observation which appears a maximum number of times. ie, the observation having the highest frequency.

Illustration 5.22

The following distribution shows the sizes of shirts sold on a textile shop in Thiruvananthapuram on a month. Calculate the mode.

Size (in inches)	:	*36*	*38*	*40*	*42*	*44*
No of shirts sold	:	*15*	*22*	*31*	*30*	*20*

Solution. In the given frequency distribution, the observation having the maximum frequency is 40. Mode is 40.

Mode from a continuous frequency distribution

$$\text{Mode} = l + \frac{(f_1 - f_0)c}{(f_1 - f_0) + (f_1 - f_2)}$$

$$\text{ie. Mode} = l + \frac{(f_1 - f_0)c}{2f_1 - f_0 - f_2}$$

Where l - lower bound of the modal class.

f_1 - frequency of the modal class.

f_0 - frequency of the preceding class to the modal class.

f_2 - frequency of the succeeding class to the modal class.

c - class interval of the modal class.

For his research on the living standards, a researcher conducted a survey on 100 persons. The distribution of the ages of the persons is attached below.

Age	:	*0-10*	*10-20*	*20-30*	*30-40*	*40-50*	*50-60*
Number of People	:	*12*	*18*	*27*	*20*	*17*	*6*

Determine the mode of this distribution.

Solution. The highest frequency = 27. ∴ The modal class is 20-30.

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

Where $l = 20, c = 10, f_0 = 18, f_1 = 27$ and $f_2 = 20$.

$$\begin{aligned} \therefore \text{Mode} &= 20 + \frac{(27-18)\times 10}{2\times 27 - 18 - 20} \\ &= 20 + \frac{9\times 10}{16} \\ &= 25.625 \end{aligned}$$

So Mode is 25.625.

Illustration 5.24

The production per day of a company (in Tons) on 60 days are given below. Calculate the mode.

Production per day	:	*21-22*	*23-24*	*25-26*	*27-28*	*29-30*
Number of days	:	*7*	*13*	*22*	*10*	*8*

Solution. Here the classes are of inclusive type. We have to convert into exclusive class before determining the mode.

Class bounds	Number of days
20.5-22.5	7
22.5-24.5	13
24.5-26.5	22
26.5-28.5	10
28.5-30.5	8

The maximum frequency is 20 so the modal class is 24.5-26.5

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

Where $l = 24.5, c = 2, f_0 = 13, f_1 = 22, f_2 = 10$

$$\text{Mode} = 24.5 + \frac{(22-13)\times 2}{2\times 22 - 13 - 10}$$
$$= 25.36$$

So mode $= 25.36$ tons.

5.3.1 Graphical location of Mode

Mode from Histogram

Like median, mode can also be located using graph. For locating mode, we use the Histogram. The steps involved in obtaining mode from histogram are

Step 1: Draw a histogram to the given data.

Step 2: Locate modal class (highest bar of histogram).

Step 3: Join diagonally the upper end points of the highest bar to the end points of the adjacent bars.

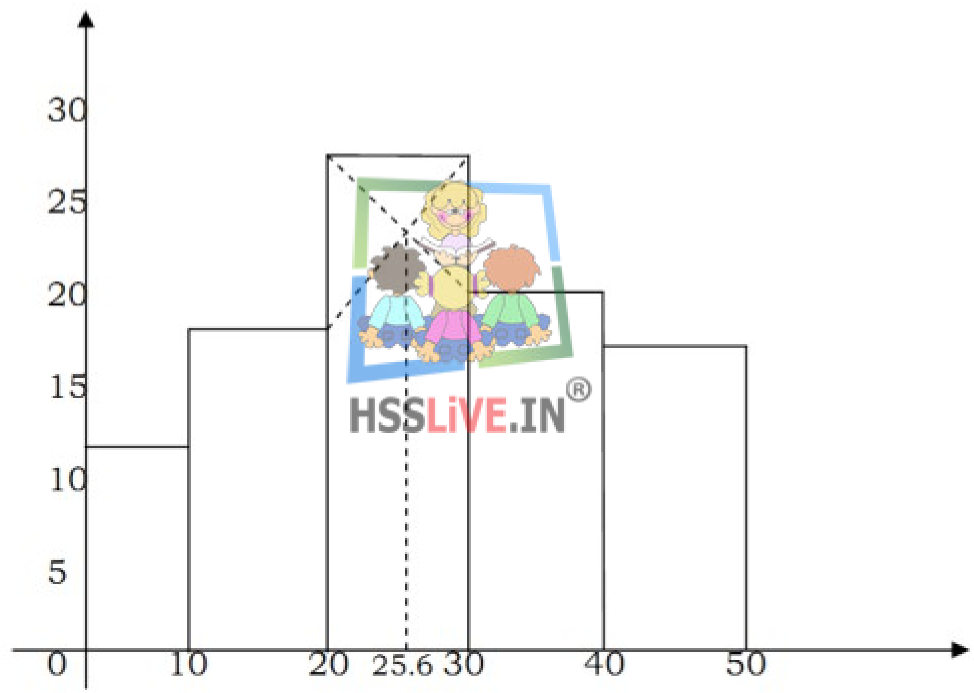
Step 4: Mark the point of intersection of the diagonals.

Step 5: Draw perpendicular from this point of intersection to the x-axis.

Step 6: The point where the perpendicular meets the x-axis gives the modal value.

Determine the mode graphically using the data provided in Illustration 5.23

Solution. To locate mode, we have to first draw the histogram.



Mode = 25.6

5.3.2 Merits and demerits of Mode

Like the other measures of central tendency, mode has its own merits and demerits. Mode is widely used in industry.

Merits

1. Mode is easy to calculate and understand. It can sometimes located merely by inspection.

2. Mode is not affected by extreme values.

3. It can be determined for open end classes.

4. Mode is the only average that works with categorical data.

Demerits

1. Mode is sometimes ill-defined. It is not defined rigidly. There are distributions with no mode, one mode, two modes etc.

2. It is not based on all the observations.

3. It is not capable of further mathematical treatments.

4. It is affected by the fluctuations of sampling.

Comparitive table-Mean, Median and Mode

Sl No	Mean	Median	Mode
1.	Defined as the arithmetic average of all the observations.	Defined as the middle value in the data set arranged in ascending or descending order.	Defined as the most frequently occurring value in the data set.
2.	Depend on all the observations.	Doesn't depend on all the observations.	Doesn't depend on all the observations.
3.	Uniquely and comprehensively defined	Can't determine in all the conditions.	Not uniquely defined for multimodal situations.
4.	Affected by extreme values.	Not affected by extreme values.	Not affected by extreme values.

5.	Can be treated mathematically. ie, means of several groups can be combined.	Can't be treated mathematically. ie, medians of several groups can't be combined.	Can't be treated mathematically. ie, modes of several groups can't be combined.
6.	More useful when data is cardinal.	More useful when the data is ordinal.	More useful when the data is nominal.

Empirical relationship among Mean, Median and Mode

For a distribution the empirical relation among mean, median and mode, given by Karl Pearson is,

Mean - Mode = 3(Mean-Median)

OR

Mode = 3 Median-2 Mean

From a partially destroyed data, it was obtained that the mode of the distribution is 63 and median is 77. Calculate the mean of the data.

Solution. We have the empirical relation,

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

$$\text{Mean} - 63 = 3(\text{Mean} - 77)$$

$$2\text{Mean} = 168$$

$$\text{Mean} = 84$$

Geometric Mean(GM)

The geometric mean of n items $x_1, x_2, x_3, \ldots, x_n$ is given by

$$GM = \sqrt[n]{x_1 \cdot x_2 \cdot x_3 \cdots x_n} = (x_1 \cdot x_2 \cdot x_3 \cdots x_n)^{\frac{1}{n}}$$

The harmonic Mean of a raw data is given by

$$HM = \frac{n}{\sum \frac{1}{x}}$$

Relation among Arithmetic Mean, Geometric Mean and Harmonic Mean

The AM, GM and HM are called the mathematical averages. The relations among them are,

1. $AM \geq GM \geq HM$

   When all the observations are the same, then $AM = GM = HM$.

2. $(GM)^2 = AM \times HM$.

   ie, $GM = \sqrt{AM \times HM}$

Partition values- Quartiles, Deciles and Percentiles

Calculation of Quartiles

(i) Quartiles from a raw data

Let there be 'n' observations. Arrange them in ascending order of magnitude. Then,

$$Q_1 = \text{value of } \left(\frac{n+1}{4}\right)^{\text{th}} \text{ item in the series}$$

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

(ii) Quartiles from a discrete frequency distribution

For computing quartiles, prepare the less than frequency table. Let N be the total frequency. Then,

$$Q_1 = \text{observation having cumulative frequency } \frac{(N+1)}{4}$$

$$Q_3 = \text{observation having cumulative frequency } \frac{3(N+1)}{4}$$

(iii) Quartiles from a continuous frequency distribution

Prepare the cumulative frequency table. Let N be the total frequency. Locate the classes having cumulative frequencies $\frac{N}{4}$ and $\frac{3N}{4}$. These classes are called the quartile classes. Then Q_1 and Q_3 are given by the formula,

$$Q_1 = l_1 + \frac{\left(\frac{N}{4} - m_1\right)c_1}{f_1}$$

$$Q_3 = l_3 + \frac{\left(\frac{3N}{4} - m_3\right)c_3}{f_3}$$

Where l_1 and l_3 are the lower bounds of quartile classes

f_1 and f_3 are the frequencies of the quartile classes

c_1 and c_3 are the class intervals of the quartile classes

m_1 and m_3 are the cumulative frequencies preceding the quartile classes

Illustration 5.34

The marks of 80 students in an examination are given below. Calculate the lower and upper quartiles.

Marks	:	*0-10*	*10 -20*	*20-40*	*40-60*	*60-80*	*80-100*
No of students	:	*8*	*10*	*22*	*25*	*10*	*5*

Solution. Let us prepare the following cumulative frequency table

Marks	No of students f	Cumulative frequency
0-10	8	8
10-20	10	18
20-40	22	40
40-60	25	65
60-80	10	75
80-100	5	80
	$N = 80$	

Here $N = 80$, $\frac{N}{4} = 20$ and $\frac{3N}{4} = 60$. Therefore, the quartile classes are 20-40 and 40-60. Then

$$Q_1 = l_1 + \frac{\left(\frac{N}{4} - m_1\right)c_1}{f_1}$$

where $l_1 = 20, c_1 = 20, f_1 = 22$ and $m_1 = 18$. So

$$Q_1 = 20 + \frac{(20 - 18) \times 20}{22}$$
$$= 20 + \frac{2 \times 20}{22}$$
$$= 21.8$$

Also

$$Q_3 = l_3 + \frac{\left(\frac{3N}{4} - m_3\right) c_3}{f_3}$$

where $l_3 = 40, c_3 = 20, f_3 = 25$, and $m_3 = 40$. So

$$Q_3 = 40 + \frac{(60 - 40) \times 20}{25}$$
$$= 20 + \frac{20 \times 20}{25}$$
$$= 56$$

Hence $Q_1 = 21.8$ and $Q_3 = 56$.

Deciles and Percentiles

Deciles are those values which divide a distribution into ten equal parts. There are 9 deciles.

Percentiles are those values which divide a distribution into hundred equal parts. There are 99 percentiles.

Median is the 5th decile and 50th percentile.

Box plot

The box plot of a data is a graphical representation based on its quartiles, as well as its smallest and largest values. It attempts to provide a visual shape of the data distribution. The box plot is also referred to as Box and Whisker plot or Box and Whisker diagram.

A box plot is a graph of a data set that consists of a line extending from the minimum value to the maximum value and a box with lines drawn at the first quartile Q 1 , the median, and the third quartile Q 3 . The lines extending from the box are called whiskers. The perpendicular line in the box is the median.

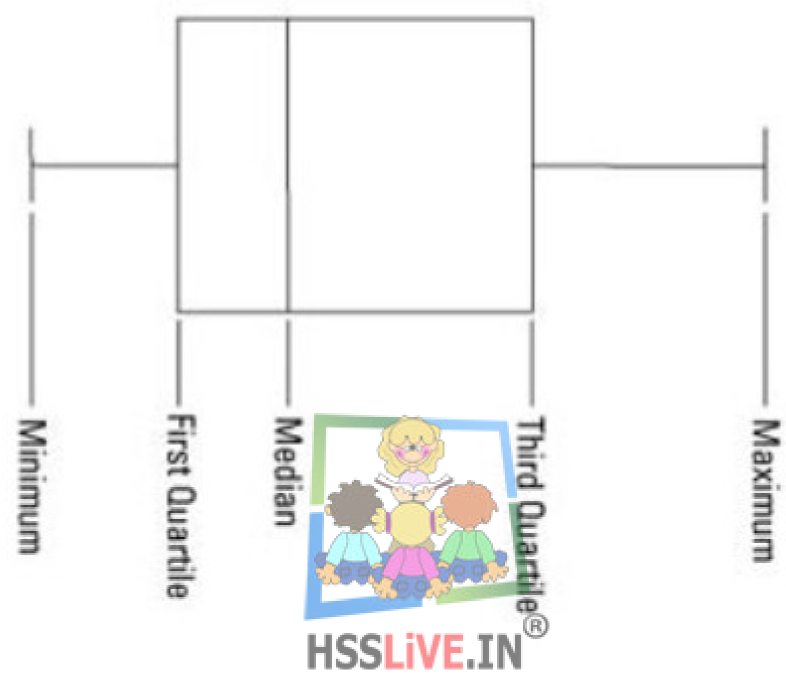


Illustration 5.35

Eleven secretaries were given a test and the scores obtained by them are as follows, 8, 7, 6, 9, 1, 3, 10, 3, 8, 4 and 7. Represent the data using a box plot.

Solution. Arranging the data in ascending order,

$$1, 3, 3, 4, 6, 7, 7, 8, 7, 9, 10$$

Number of observations, $n = 11$. To draw the box plot, we have to determine the following.

$$\begin{aligned}
\text{Minimum value} &= 1 \\
Q_1 &= \frac{n+1}{4}^{\text{th}} \text{ item} \\
&= \frac{11+1}{4}^{\text{th}} \text{ item} \\
&= 3^{\text{rd}} \text{ item} \\
&= 3 \\
\text{Median} &= \frac{n+1}{2}^{\text{th}} \text{ item} \\
&= 6^{\text{th}} \text{ item} \\
&= 7 \\
Q_3 &= \frac{3(n+1)}{4}^{\text{th}} \text{ item} \\
&= 9^{\text{th}} \text{ item} \\
&= 8 \\
\text{Maximum value} &= 10
\end{aligned}$$