2. Measures of Central Tendency:

Summarization of data is a necessary function of any statistical analysis. The data is summarized in the form of tables and frequency distributions. In order to bring the characteristics of the data, these tables and frequency distributions need to be summarized further. A measure of central tendency or an average is very essential and an important summary measure in any statistical analysis.

An average is a single value which can be taken as a representative of the whole distribution. There are five types of measures of central tendency or averages which are commonly used.

(i) Arithmetic mean

(ii) Median

(iii) Mode

(iv) Geometric mean

(v) Harmonic mean

A good measure of average must have the following characteristics:

(i) It should be rigidly defined so that different persons obtain the same value for a given set of data.

(ii) It should be easy to understand and easy to calculate.

(iii) It should be based on all the observations of the data.

(iv) It should be easily subjected to further mathematical calculations.

(v) It should not be much affected by the fluctuations of sampling.

(vi) It should not be unduly affected by extreme observations.

(vii) It should be easy to interpret.

2.1. Arithmetic Mean :

The arithmetic mean of a set of observations is their sum divided by the number of observations. Let x1, x2,..., xn be n observations. Then their average or arithmetic mean is given by

$$\overline{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{\sum x}{n}$$

For example, the marks obtained by 10 students in Class XII in a physics examination are 25, 30, 21, 55, 40, 45, 17, 48, 35, 42. The arithmetic mean of the marks is given by

$$\overline{x} = \frac{\sum x}{n} = \frac{25+30+21+55+40+45+17+48+35+42}{10} = \frac{358}{10} = 35.8$$

If n observations consist of n distinct values denoted by x1, x2, ..., xn of the observed variable x occurring with frequencies f1, f2, ..., fn respectively then the arithmetic mean is given by

$$\overline{x} = \frac{f_1 x_1 + f_2 x_2 + \cdots + f_n x_n}{f_1 + f_2 + \cdots + f_n} = \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i} = \frac{\sum_{i=1}^{n} f_i x_i}{N} = \frac{\sum fx}{N}$$

- **Arithmetic Mean of Grouped data**

In case of grouped or continuous frequency distribution the arithmetic mean is given by

$$\bar{x} = \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i} = \frac{\sum fx}{N}, \text{ where } N = \sum_{i=1}^{n} f_i$$

and x is taken as the midvalue of the corresponding class.

Example 1:-

Find the arithmetic mean from the following frequency distribution:

x	5	6	7	8	9	10	11	12	13	14
f	25	45	90	165	112	96	81	26	18	12

Solution

x	f	fx
5	25	125
6	45	270
7	90	630
8	165	1320
9	112	1008
10	96	960
11	81	891
12	26	312
13	18	234
14	12	168
	$\sum f = 670$	$\sum fx = 5918$

$$N = \sum f = 670$$

$$\bar{x} = \frac{\sum fx}{N} = \frac{5918}{670} = 8.83$$

Example 2:

.

Find the arithmetic mean of the marks from the following data:

Marks	0–10	10–20	20–30	30–40	40–50	50–60
Number of students	12	18	27	20	15	8

Solution

Marks	Number of students (*f*)	Midvalue (*x*)	*fx*
0–10	12	5	60
10–20	18	15	270
20–30	27	25	675
30–40	20	35	700
40–50	15	45	675
50–60	8	55	440
	$\sum f = 100$		$\sum fx = 2820$

$$N = \sum f = 100$$

$$\bar{x} = \frac{\sum fx}{N} = \frac{2820}{100} = 28.20$$

- **Properties of arithmetic Mean**

1. The algebraic sum of deviations from the mean is zero. If the mean of n observations $x_1, x_2, ..., x_n$ is $\bar{x}$ then $(x_1 - \bar{x}) + (x_2 - \bar{x}) + \cdots + (x_n - \bar{x}) = 0$, i.e., $\sum (x - \bar{x}) = 0$.
2. The sum of squares of the deviations is minimum when taken about the mean.
3. If $\bar{x}_1, \bar{x}_2, ..., \bar{x}_k$ are the means of k series of sizes $n_1, n_2, ..., n_k$ respectively then the mean $\bar{x}$ of the composite series is given by

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

$$= \frac{\sum_{i=1}^{k} n_i x_i}{\sum_{i=1}^{k} n_i}$$

Example 3 :

Find the value of p for the following distribution whose mean is 11.37.

x	5	7	p	11	13	16	20
f	2	4	29	54	11	8	4

Solution

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

$$11.37 = \frac{(5\times 2)+(7\times 4)+29p+(11\times 54)+(13\times 11)+(16\times 8)+(20\times 4)}{2+4+29+54+11+8+4}$$

$$= \frac{10+28+29p+594+143+128+80}{112}$$

$$= \frac{983+29p}{112}$$

$$\therefore \;\; p = 10.015 \approx 10$$

2.2 Median :

Median is the central value of the variable when the values are arranged in ascending or descending order of magnitude. It divides the distribution into two equal parts.

When the observations are arranged in the order of their size, median is the value of that item which has equal number of observations on either side.

In case of ungrouped data, if the number of observations is odd then the median is the middle value after the values have been arranged in ascending or descending order of magnitude. If the number of observations is even, there are two middle terms and the median is obtained by taking the arithmetic mean of the middle terms.

Examples 3 :

(i) The median of the values 20, 15, 25, 28, 18, 16, 30, i.e., 15, 16, 18, 20, 25, 28, 30 is 20 because n = 7, i.e., odd and the median is the middle value, i.e., 20.

(ii) The median of the values 8, 20, 50, 25, 15, 30, i.e., 8, 15, 20, 25, 30, 50 is the arithmetic mean of the middle terms, i.e., $\frac{20+25}{2} = 22.5$ because n = 6, i.e., even.

In case of **discrete frequency distribution**, the median is obtained by considering the cumulative frequencies. The steps for calculating the median are given below:

(i) Arrange the values of the variables in ascending or descending order of magnitudes.

(ii) Find $\frac{N}{2}$ where N = $= \sum f$

(iii) Find the cumulative frequency just greater than $\frac{N}{2}$ and determine the corresponding value of the variable.

(iv) The corresponding value of x is the median.

Example 4 :

The following table represents the marks obtained by a batch of 12 students in certain class tests in physics and chemistry.

Marks (Physics)	53	54	32	30	60	46	28	25	48	72	33	65
Marks (Chemistry)	55	41	48	49	27	25	23	20	28	60	43	67

Indicate the subject in which the level of achievement is higher.

Solution

The level of achievement is higher in that subject for which the median marks are more. Arranging the marks in two subjects in ascending order,

Marks (Physics)	25	28	30	32	33	46	48	53	54	60	65	72
Marks (Chemistry)	20	23	25	27	28	41	43	48	49	55	60	67

Since the number of students is 12, the median is the arithmetic mean of the middle terms.

$$\text{Median marks in Physics} = \frac{46+48}{2} = 47$$

$$\text{Median marks in Chemistry} = \frac{41+43}{2} = 42$$

Since the median marks in physics are greater than the median marks in chemistry, the level of achievement is higher in physics.

Example 5:

Obtain the median for the following frequency distribution.

x	0	1	2	3	4	5	6	7
f	7	14	18	36	51	54	52	18

Solution

x	f	Cumulative Frequency
0	7	7
1	14	21
2	18	39
3	36	75
4	51	126
5	54	180
6	52	232
7	18	250

$$N = 250$$

$$\frac{N}{2} = \frac{250}{2} = 125$$

The cumulative frequency just greater than $\frac{N}{2}$ = 125 is 126 and the value of x corresponding to 126 is 4. Hence, the median is 4.

- **Median for continuous frequency distribution**

In case of continuous frequency distribution (less than frequency distribution), the class corresponding to the cumulative frequency just greater than , $\frac{N}{2}$ is called the median class, and the value of the median is given by

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

where l is the lower limit of the median class
f is the frequency of the median class
h is the width of the median class
c is the cumulative frequency of the class preceding the median class
N is sum of frequencies, i.e., $N = \sum f$

In case of 'more than' or 'greater than' type of frequency distributions, the value of the median is given by

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

where u is the upper limit of the median class
f is the frequency of the median class
h is the width of the median class
c is the cumulative frequency of the class succeeding the median class

Example 6:

The following table gives the weekly expenditures of 100 workers. Find the median weekly expenditure.

Weekly Expenditure (in ₹)	0–10	10–20	20–30	30–40	40–50
Number of workers	14	23	27	21	15

Solution

Weekly Expenditure (in ₹)	Number of Workers i.e., frequency (f)	Cumulative Frequency
0–10	14	14
10–20	23	37
20–30	27	64
30–40	21	85
40–50	15	100

$$N = 100$$

$$\frac{N}{2} = \frac{100}{2} = 50$$

The cumulative frequency just greater than $\frac{N}{2} = 50$ is 64 and the corresponding class 20–30 is the median class.

Here, $\frac{N}{2} = 50, \quad l = 20, \quad h = 10, \quad f = 27, \quad c = 37$

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

$$= 20 + \frac{10}{27}(50 - 37)$$

$$= 24.815$$

Example 7 :

From the following data, calculate the median:

Marks (Less than)	5	10	15	20	25	30	35	40	45
No. of Students	29	224	465	582	634	644	650	653	655

Solution

This is a ‘less than’ type of frequency distribution. This will be first converted into class intervals.

Class Intervals	Frequency	Less than CF
0–5	29	29
5–10	195	224
10–15	241	465
15–20	117	582
20–25	52	634
25–30	10	644
30–35	6	650
35–40	3	653
40–45	2	655

$N = 655$

Since $\frac{N}{2} = \frac{655}{2} = 327.5$, the median class is 10–15.

Here, $l = 10, \quad h = 5, \quad f = 241, \quad c = 224$

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

$$= 10 + \frac{5}{241}(327.5 - 224)$$

$$= 12.147$$

2.3. Mode:

Mode is the value which occurs most frequently in a set of observations and around which the other items of the set are heavily distributed. In other words, mode is the value of the variable which is most frequent or predominant in the series. In case of a discrete frequency distribution, mode is the value of x corresponding to the maximum frequency.

Examples 8 :

(i) In the series 6, 5, 3, 4, 3, 7, 8, 5, 9, 5, 4, the value 5 occurs most frequently. Hence, the mode is 5.

(ii) Consider the following frequency distribution:

x	1	2	3	4	5	6	7	8
f	4	9	16	25	22	15	7	3

The value of x corresponding to the maximum frequency, viz., 25, is 4. Hence, the mode is 4.

Note: For an asymmetrical frequency distribution, the difference between the mean and the mode is approximately three times the difference between the mean and the median.

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

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

This is known as the empirical formula for calculation of the mode.

- **Mode for a continuous frequency distribution:**

In case of a continuous frequency distribution, the class in which the mode lies is called the modal class and the value of the mode is given by

$$\text{Mode} = l + h\left(\frac{f_m - f_1}{2f_m - f_1 - f_2}\right)$$

where L is the lower limit of the modal class

h is the width of the modal class

F_m is the frequency of the modal class

f1 is the frequency of the class preceding the modal class

f2 is the frequency of the class succeeding the modal class

This method of finding mode is called the method of interpolation. This formula is applicable only to a unimodal frequency distribution.

Example 9 :

Find the mode for the following data:

Profit per shop	0–100	100–200	200–300	300–400	400–500	500–600
No. of Shops	12	18	27	20	17	6

Solution

Since the maximum frequency is 27 which lies in the class 200–300, the modal class is 200–300.

Here, $l = 200$, $h = 100$, $f_m = 27$, $f_1 = 18$, $f_2 = 20$

$$\text{Mode} = l + h\left(\frac{f_m - f_1}{2f_m - f_1 - f_2}\right)$$

$$= 200 + 100\left[\frac{27 - 18}{2(27) - 18 - 20}\right]$$

$$= 256.25$$

Example 10:

Find the mode for the following distribution:

Class intervals	0–10	10–20	20–30	30–40	40–50
Frequency	45	20	14	7	3

Solution

Since the highest frequency occurs in the first-class interval, the interpolation formula is not applicable. Thus, empirical formula is used for calculation of mode.

Class intervals	Frequency	CF	Midvalue	$d = \frac{x - 25}{10}$	fd
0–10	45	45	5	–2	–90
10–20	20	65	15	–1	–20
20–30	14	79	25	0	0
30–40	7	86	35	1	7
40–50	3	89	45	2	6
	$\sum f = 89$				$\sum fd = -97$

$$N = \sum f = 89$$

Since $\frac{N}{2} = \frac{89}{2} = 44.5,$ the median class is 0–10.

Here, $l = 0, \quad h = 10, \quad f = 45, \quad c = 0$

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

$$= 0 + \frac{10}{45}(44.5 - 0)$$

$$= 9.89$$

$$\text{Mean} = a + h\,\frac{\sum fd}{N}$$

$$= 25 + 10\left(\frac{-97}{89}\right)$$

$$= 14.1$$

$$\text{Hence, mode} = 3\,\text{Median} - 2\,\text{Mean}$$

$$= 3(9.89) - 2(14.1)$$

$$= 1.47$$

Example 11:

The following table gives the incomplete income distribution of 300 workers of a company, where the frequencies of the classes 3000–4000 and 5000–6000 are missing. If the mode of the distribution is ` 4428.57, find the missing frequencies.

Monthly Income (in ₹)	No. of Workers
1000–2000	30
2000–3000	35
3000–4000	?
4000–5000	75
5000–6000	?
6000–7000	30
7000–8000	15

Solution

Let f1 and f2 be the frequencies of the classes 3000–4000 and 5000–6000 respectively. f1 + f2 = 300 – (30 + 35 + 75 + 30 + 15) = 115

Since the mode is 4428.57, the modal class is 4000–5000.

Here, l = 4000, h = 1000, fm = 75

$$\text{Mode} = l + h\left(\frac{f_m - f_1}{2f_m - f_1 - f_2}\right)$$

$$= l + h\left[\frac{f_m - f_1}{2f_m - (f_1 + f_2)}\right]$$

$$4428.57 = 4000 + 1000\left[\frac{75 - f_1}{2(75) - 115}\right]$$

$$\therefore \quad f_1 = 60$$

$$f_2 = 115 - 60 = 55$$

Homework 1:

Find the mean of the following marks obtained by students of a class:

Marks	15	20	25	30	35	40
No. of Students	9	7	12	14	15	6

Calculate the mean for the following data:

Heights (in cm)	135–140	140–145	145–150	150–155	155–160	160–165	165–170	170–175
No. of boys	4	9	18	28	24	10	5	2

Calculate the median of the following data:

x	3–4	4–5	5–6	6–7	7–8	8–9	9–10	10–11
f	3	7	12	16	22	20	13	7

[Ans.: 7.55]

The weekly wages of 1000 workers of a factory are shown in the following table:

Weekly wages (less than)	425	475	525	575	625	675	725	775	825	875
No. of Workers	2	10	43	123	293	506	719	864	955	1000

[Ans.: 673.59]

Calculate the mode of the following distribution:

x	0–5	5–10	10–15	15–20	20–25	25–30	30–35	35–40	40–45
f	20	24	32	28	20	16	37	10	18

[Ans.: 13.33]