Measures of Central Tendency

The primary purpose of statistical methods is to summerise the information contained in any set of collected data. The purpose is served by classifying the data in form of a frequency distribution and using various graphs, viz., Line diagrams, Bar diagrams, Pictorial diagrams, Representation of percentages, Statistical maps. When the data relate to a variable, the process of summerisation can be taken a long step further by using certain descriptive measures. The aim is to focus on certain features of the data which will describe the general nature of the data. The two most important features are central tendency and dispersion.

Central Tendency:

Let us consider the following table.

Table 1: Yield per plant for 12 tomato plants of a particular variety

Plant No.	Yield (gm.)	Plant No.	Yield (gm.)
1.	1,216	7.	1,202
2.	1,374	8.	1,372
3.	1,167	9.	1,278
4.	1,232	10.	1,141
5.	1,407	11.	1,221
6.	1,453	12.	1,329

From Table 1, it is clearly evident that the figures seem to cluster around some point between 1,200 gm and 1,300 gm. However, we need a single value, the central value, to represent the whole set of figures. Such a representative or typical value of a variable is called the measure of central tendency or an average.

Three comonly used measures of central tendency are

(i) Arithmatic mean

(ii) Median

(iii) Mode

Arithmatic Mean :

Let us denote the variable by x and the corresponding n values of the ~~vad~~ variable x by $x_1, x_2, \ldots \ldots x_n$.

For example, let x represents the height of n students and the corresponding heights are represented by $x_1, x_2, \ldots, x_n$.

Then the arithmatic mean (a.m) of x is given by $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$ ——— (1.1)

Example 1. For ~~the~~ the data given in Table 1. the a.m is $\frac{1,216 + 1,374 + \cdots + 1,329}{12}$

$= \frac{15,392}{12}$

$= 1,282.67$ gm.

Example 2.

Let us consider the following table

Table 2. Frequency distribution of no. of peas per pod for 198 pods.

No. of peas	Frequency (f_i)
1	4
2	33
3	76
4	50
5	26
6	8
7	1
Total	$198 = \sum_{i=1}^{7} f_i$

In the above table, the value 1 occurs 4 times, value 2 occurs 33 times, and so on.

Therefore, if in the above example x represents the no. of peas per pod and the corresponding the value of x, i.e., x_i $(i=1,2,\ldots 198)$ represents the no. of peas in ith pod, then

$$\sum_{i=1}^{198} x_i = 1\times4 + 2\times33 + 3\times76 + 4\times50 + 5\times26 + 6\times8 + 7\times1$$

$$= 683.$$

Hence, a.m of x is given by

$$\bar{x} = \frac{1\times4 + 2\times33 + 3\times76 + 4\times50 + 5\times26 + 6\times8 + 7\times1}{198}$$

$$= \frac{683}{198} = 20.697$$

Therefore, if the numbers $x_1, x_2, \ldots, x_n$ occur $f_1, f_2, \ldots, f_n$ times, respectively (i.e., occur with frequencies $f_1, f_2, \ldots, f_n$) the arithmatic mean is given by

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i f_i}{\sum_{i=1}^{n} f_i} = \frac{1}{N}\sum_{i=1}^{n} x_i f_i \quad \text{——} (1.2)$$

where $N = \sum_{i=1}^{n} f_i$ is the total frequency.

When we have a frequency table, ~~where the frequencies~~ which represents the frequencies in the different classes, then also we will use the formula (1.2) for calculating the arithmatic mean. However, in this case x_i will represent the mid value of ith class interval. But in this case (1.2) will give only an approximate value of the mean. The error of approximation will be negligible ~~class~~ provided the range of x is very large compared to the width of the class-intervals.

Example 3.

(a) Find the arithmatic mean of the following frequency distribution

x_i:	1	2	3	4	5	6	7
f_i:	5	9	12	17	14	10	6

(b) Calculate the arithmatic mean of the marks from the following table :

Marks :	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50
No. of students	12	18	27	20	17

Marks	50 – 60
No. of students	6

Soln.

(a)

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

$$\therefore \quad \bar{x} = \frac{\Sigma x_i f_i}{\Sigma f_i} = \frac{299}{73} = 4.0959$$

(b)

Marks	Mid point (x_i)	No. of students (f_i)	$x_i f_i$
0 – 10	5	12	60
10 – 20	15	18	270
20 – 30	25	27	675
30 – 40	35	20	700
40 – 50	45	17	765
50 – 60	55	6	330
		100	2800

$$\therefore \text{Arithmatic Mean (A.M)} = \frac{\Sigma x_i f_i}{\Sigma f_i}$$

$$= \frac{2800}{100} = 28.$$

$\therefore$ Average marks of the students is 28.

It may be noted here that if the values of x_i's (and) or f_i's, the calculation of mean by formula (1.2) is quite time-consuming and tedious.

Let $d_i = x_i - A$, $i = 1, 2, \cdots, n$

$\Rightarrow f_i d_i = f_i (x_i - A) = f_i x_i - A f_i$, $i = 1, 2, \cdots, n$

$\Rightarrow \frac{1}{N} \sum f_i d_i = \frac{1}{N} \sum_{i=1}^{n} x_i f_i - A$ $\quad \because N = \sum_{i=1}^{n} f_i$

$= \bar{x} - A.$

$$\Rightarrow \boxed{\bar{x} = A + \frac{1}{N} \sum_{i=1}^{n} f_i d_i} \quad \text{———} (1.3)$$

A is any arbitrary point.

Let us consider grouped or continuous frequency distribution.

Let $d_i = \dfrac{x_i - A}{h}$, $i = 1, 2, \cdots, n$

A is an arbitrary point

h is the common magnitude of class interval

Now $h d_i = x_i - A$

$\Rightarrow \sum_{i=1}^{n} h d_i f_i = \sum_{i=1}^{n} x_i f_i - A \sum_{i=1}^{n} f_i$

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

$$\Rightarrow \boxed{\bar{x} = A + \frac{h}{N} \sum_{i=1}^{n} d_i f_i} \quad \text{———} (1.4)$$

Example 4

Calculate the simple mean/arithmatic mean/mean for the following frequency distribution.

Class interval	0 – 8	8 – 16	16 – 24	24 – 32	32 – 40	40–4
Frequency	8	7	16	24	15	7

Soln.

Class interval	Mid Value (x_i)	Frequency (f_i)	$d_i = \frac{x_i - A}{h}$	f_i
0 – 8	4	8	–3	–2
8 – 16	12	7	–2	–14
16 – 24	20	16	–1	–16
24 – 32	28 (= A)	24	0	0
32 – 40	36	15	1	15
40 – 48	44	7	2	14
Total		77 (= N)		–25

Let us consider $A = 28$

& $h = 8$

$$\therefore \quad \bar{x} = A + \frac{h}{N} \sum f_i d_i = 28 + \frac{8}{77}(-25)$$

$$= 25.404$$

The weighted Arithmatic Mean:

In calculating arithmatic mean we assume that all th items in the distribution have equal importance. But in practice this may not be so. If some items in a distribution are more important than others, then this point must be considered in calculating the average. In such cases, proper weights must be give to various items. The weights attached to each item being proportional to the importance of the item in the distribution.

For example, let w_i be the weight attached to the item x_i ; $i = 1, 2, \ldots, n$. Then we define:

Weighted arithmatic mean (or weighted mean)

$$= \sum_{i=1}^{n} w_i x_i \Big/ \sum_{i=1}^{n} w_i \qquad (1.5)$$

Example 5

If a final examination in a course is weighted 3 times as much as a quiz and a student has a final examination grade of 85 and quiz grades of 70 and 90, then the mean grade is

$$\bar{x} = \frac{1 \times 70 + 1 \times 90 + 3 \times 85}{1 + 1 + 3} = \frac{415}{5} = 83.$$

Median :

Median of a distibution is the value of the ~~valu~~ variables which ~~defi~~ devides the entire set of values into two equal parts. That is ~~its~~ median is the value such that the no. of observation above it is equal to the no. of observations below it. The median is thus a positional average.

In case of ungrouped data, if the no. of observations is add then median is the middle value after the values have been arranged in ascending or descending order of magnitude. In case of even no. of observations, there are two middle terms and median is obtained by taking the arithmatic mean of the middle terms.

For example, the median of the ~~terms~~ values 25, 20, 15, 10, 5, 21, 7, i.e., 5, 7, 10, 15, 20, 21, 25 is 15. and the median of 8, 16, 12, 1, 2, 9, 15, 30, 25, 4, i.e, 1, 2, 4, 8, 9, 12, 15, 16, 25, 30 is $\frac{1}{2}(9+12) = 10.5$

Remark: In case of even no. of observations, in fact any value lying between the two middle values can be taken as median but conventionally we take it to be the mean of the middle term.

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

(i) Find $\frac{1}{2}N$, where $N = \sum_{i=1}^{n} f_i$

(ii) See the (less than) cumulative frequency (c.f.) just greater than $\frac{1}{2}N$.

(iii) The corresponding value of x is median.

Example 6.

Obtain the median for the following frequency distribution:

x_i :	1	2	3	4	5	6	7	8	9
f_i :	8	10	11	16	20	25	15	9	6

Soln.

x_i	f_i	c.f. (less than type)
1	8	8
2	10	18
3	11	29
4	16	45
$\boxed{5}$	20	$\boxed{65}$
6	25	90
7	15	105
8	9	114
9	6	120 (= N)

$\therefore \frac{1}{2}N = \frac{120}{2} = 60$

$\therefore$ The (less than type) cumulative frequency just greater than $N/2$ is 65 and the corresponding value of x_i is 5.

$\therefore$ Median is 5.

Median for continuous frequency distribution

In case of continuous frequency distribution, the class corresponding to the (less than) cumulative frequency just greater than $\frac{1}{2}N$ is called the median class and the value of median is obtained by the following formula

$$\text{Median} = l + \frac{h}{f}\left(\frac{N}{2} - c\right) \quad \text{———} (1.6)$$

where $l \rightarrow$ is the lower limit of the median class

$f \rightarrow$ is the frequency of the median class

$h \rightarrow$ is the length of the median class

$c \rightarrow$ is the c.f. of the class preceeding the median class.

Example 7

Find the median wage of the following distribution.

Wages (in Rs.):	2000-3000	3000-4000	4000-5000	5000-6000
No. of workers	3	5	20	10

Wages (in Rs.):	6000 - 7000
No. of workers:	5

Soln.

Wages (in Rs.)	No. of workers	c.f
2000 - 3000	3	3
3000 - 4000	5	8
4000 - 5000	20	28 ← Median class
5000 - 6000	10	38
6000 - 7000	5	43

$N = 43 \Rightarrow \frac{N}{2} = 21.5$

Cumulative frequency just greater than 21.5 is 28 and the corresponding class is 4000 - 5000.

Thus the median class is 4000 - 5000.

Hence, Median $= 4000 + \frac{1000}{20}(21.5 - 8)$

$= 4675$

Thus median wage is Rs. 4,675

Mode :

Mode is the value which occurs most frequently in a set of observations. In other words, mode is the value of the variable which is predominant in the series.

For example, in the following frequency distribution

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

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

Made for continuous frequency distribution:

In case of continuous frequency distribution, made is given by the formula

$$\text{Made} = l + \frac{h(f_1 - f_0)}{(f_1 - f_0) - (f_2 - f_1)}$$

$$= l + \frac{h(f_1 - f_0)}{2f_1 - f_0 - f_2} \qquad \text{——— (1.7)}$$

where $l \rightarrow$ lower limit of the madal clan

$h \rightarrow$ length / magnitude of the madal clan

$f_1 \rightarrow$ frequency of the madal clan

$f_0 \rightarrow$ frequency of the clan preceeding the madal clan

$f_2 \rightarrow$ frequency of the clan succeeding the madal clan.

& the madal clan is the clan with maximum frequency.

Example 8.

Find the made for the following distribution

Clan-interval :	0-10	10-20	20-30	30-40	40-50
Frequency :	5	8	7	12	28

Clan-interval :	50-60	60-70	70-80
Frequency :	20	10	10

Soln. Maximum frequency is 28

$\therefore$ the madal clan is 40-50

$$\therefore \text{Made} = 40 + \frac{10(28-12)}{(2\times 28 - 12 - 20)} = 40 + 6.666 = 46.67$$