A. ARITHMETIC MEAN

The most popular and widely used measure of representing the entire data by one value is what most laymen call an 'average' and what the statisticians call the arithmetic mean*. Its value is obtained by

*It should be noted that the statisticians do not like the term 'average' because it has too loose a connotation. It has different meanings, for example, an average person, an average wage, an average height, etc. It can refer to either mean, median, mode, geometric mean, harmonic mean or any other average. In practice, arithmetic mean is so popular that the word 'mean or average' alone without qualification is implied to denote this particular type of average. That is, when anyone speaks of 'the mean' or 'the average' of a series of observations, it may, as a rule, be assumed that the arithmetic mean is meant unless otherwise stated.

adding together all the items and by dividing this total by the number of items. Arithmetic mean may either be

(*i*) simple arithmetic mean, or

(*ii*) weighted arithmetic mean.

Calculation of Simple Arithmetic Mean—Individual Observations

The process of computing mean in case of individual observations (*i.e.*, where frequencies are not given) is very simple. Add together the various values of the variable and divide the total by the number of items. Symbolically :

$$\bar{X}^* = \frac{X_1+X_2+X_3+\ldots+X_n}{N}, \quad \text{or} \quad \bar{X} = \frac{\Sigma X}{N}$$

Here $\bar{X}$=Arithmetic Mean, ΣX=Sum of all the values of the variable X, *i.e.*, $X_1, X_2, X_3, \ldots, X_n$: N=Number of observations.

Steps. The formula involves two steps in calculating mean :

(*i*) Add together all the values of the variable X and obtain the total, *i.e.*, ΣX.

(*ii*) Divide this total by the number of the observations, *i.e.*, N.

Illustration 1. The following table gives the monthly income of 10 emyloyees in an office :

Income (Rs.)	1780	1760	1690	1750	1840	1920	1100	1810	1050	1950

Calculate the arithmetic mean of incomes.

Let income be denoted by the symbol X.

Solution. CALCULATION OF ARITHMETIC MEAN

Employee	*Monthly Income (Rs.)*	*Employee*	*Monthly Income (Rs.)*
1	1,780	6	1,920
2	1,760	7	1,100
3	1,690	8	1,810
4	1,750	9	1,050
5	1,840	10	1,950
		$N=10$	$\Sigma X=16,650$

$$\bar{X} = \frac{\Sigma X}{N} \qquad \Sigma X = 16,650, \qquad N = 10$$

Hence $$\bar{X} = \frac{16,650}{10} = 1,665$$

Thus the average income is Rs. 1,665

Short-cut method. The arithmetic mean can be calculated by using, what is known as an *arbitrary origin*. When deviations are taken

*The arithmetic mean of a sample is designated by the symbol $\bar{X}$, which is read 'X-bar' and the arithmetic mean of a population is designated by the Greek Letter 'μ' pronounced as "*mu*".

The symbol Σ is the letter capital sigma of the Greek alphabet and is used in mathematics to denote the sum of the values.

from an arbitrary origin, the formula for calculating arithmetic mean is

$$\bar{X}=A+\frac{\Sigma d}{N}$$

where A is the assumed mean and d is the deviation of items from assumed mean, *i.e.*, $d=(X-A^*)$.

Steps. (1) Take an assumed mean.

(2) Take the deviations of items from the assumed mean and denote these deviations by d.

(3) Obtain the sum of these deviations, *i.e.*, Σd.

(4) Apply the formula : $\bar{X}=A+\frac{\Sigma d}{N}$

From illustration 1 calculate arithmetic mean by taking 1800 as the assumed mean :

CALCULATION OF ARITHMETIC MEAN

Employee	*Income* (*Rs.*)	$(X-1800)$ d
1	1,780	− 20
2	1,760	− 40
3	1,690	−110
4	1,750	− 50
5	1,840	+ 40
6	1,920	+120
7	1,100	−700
8	1,810	+ 10
9	1,050	−750
10	1,950	+150
$N=10$		$\Sigma d=-1350$

$$\bar{X}=A+\frac{\Sigma d}{N}$$

$$A=1800,\ \Sigma d=1350,\ N=10$$

$$\bar{X}=1800-\frac{1,350}{10}=1,800-135=1,665$$

Hence the average income is Rs. 1,665.

Note. The reader will find that the calculations here are more than what we had when we used the formula

$$\bar{X}=\frac{\Sigma X}{N}.$$

This is true for ungrouped data. But for grouped data considerable saving in time is possible by adopting the short-cut method.

*Any value whether existing in the data or not can be taken as the assumed mean and the final answer would be the same. However, the nearer the assumed mean is to the actual mean, the lesser are the calculations.

Calculation of Arithmetic Mean—Discrete Series

In discrete series arithmetic mean may be computed by applying

(*i*) Direct method, or

(*ii*) Short-cut method.

Direct Method. The formula for computing mean is

$$\bar{X}=\frac{\Sigma fX}{N}$$

where, f=Frequeney ; X=The variable in question ; N*=Total number of observations, *i.e.*, Σf.

Steps : (*i*) Multiply the frequency of each row with the variable and obtain the total ΣfX.

(*ii*) Divide the total obtained by step (*i*) by the number of observations, *i.e.*, total frequency.

Illustration 2. From the following data of the marks obtained by 60 students of a class calculate the arithmetic mean :

Marks	*No. of Students*	*Marks*	*No. of Students*
20	8	50	10
30	12	60	6
40	20	70	4

Let the marks be denoted by X and the number of students by f.

Solution : CALCULATION OF ARITHMETIC MEAN

Marks X	*No. of Students* f	fX
20	8	160
30	12	360
40	20	800
50	10	500
60	6	360
70	4	280
	$N=60$	$\Sigma fX=2,460$

$$\bar{X}=\frac{\Sigma fX}{N}=\frac{2,460}{60}=41$$

Hence the average marks=41.

Short-cut Method. According to this method,

$$\bar{X}=A+\frac{\Sigma fd}{N}$$

where A=Assumed mean ; $d=(X-A)$; N=Total number of observations, *i.e.*, Σf.

---

*The reader should note carefully that in discrete and continuous frequency distributions the total number of observations, *i.e.*, N=the sum of frequency or $N=\Sigma f$.

Steps : (*i*) Take an assumed mean.

(*ii*) Take the deviations of the variable X from the assumed mean and denote the deviations by d.

(*iii*) Multiply these deviations with the respective frequency and take the total Σfd.

(*iv*) Divide the total obtained in third step by the total frequency.

Illustration 3. Calculate arithmetic mean by the short-cut method using frequency distribution of illustration 2.

Solution. CALCULATION OF ARITHMETIC MEAN

Marks X	*No. of Students* f	$(X-40)$ d	fd
20	8	−20	−160
30	12	−10	−120
40	20	0	0
50	10	+10	+100
60	6	+20	+120
70	4	+30	+120
	$N=60$		$\Sigma fd=60$

$$\bar{X}=A+\frac{\Sigma fd}{N}=40+\frac{60}{60}=40+1=41.$$

Calculation of Arithmetic Mean—Continuous Series

In continuous series, arithmetic mean may be computed by applying any of the following methods :

(*i*) Direct method.

(*ii*) Short-cut method.

Direct Method. When direct method is used

$$\bar{X}=\frac{\Sigma fm}{N}$$

where m*=mid-point of various classes ; f=the frequency of each class ; N=the total frequency.

Steps : (*i*) Obtain the mid-point of each class and denote it by m.

(*ii*) Multiply these mid-points by the respective frequency of each class and obtain the total Σfm.

(*iii*) Divide the total obtained in step (*i*) by the sum of the frequency, *i.e.*, N.

Note. For sake of clarity and ease of understanding we have used the terms individual observations, discrete series and continuous series throughout the text. However, the reader should be familiar with the terms grouped data and ungrouped data also. Ungrouped data refer to the individual observations whereas grouped data refer to the continuous series and the discrete series.

$$*\text{Mid-point}=\frac{\text{Lower limit}+\text{Upper limit}}{2}.$$

Illustration 4. (*a*) From the following data compute arithmetic mean by direct method :

Marks	0—10	10—20	20—30	30—40	40—50	50—60
No. of students	5	10	25	30	20	10

Solution. CALCULATION OF ARITHMETIC MEAN BY DIRECT METHOD

Marks	*Mid-point* m	*No. of Students* f	fm
0—10	5	5	25
10—20	15	10	150
20—30	25	25	625
30—40	35	30	1,050
40—50	45	20	900
50—60	55	10	550
		$N=100$	$\Sigma fm=3,300$

$$\bar{X}=\frac{\Sigma fm}{N}=\frac{3300}{100}=33.$$

Short-cut Method. When short-cut method is used, arithmetic mean is computed by applying the following formula :

$$\bar{X}=A+\frac{\Sigma fd}{N}$$

when A=assumed mean ; d=deviations of mid-points from assumed mean, *i.e.*, $(m-A)$; N=total number of observations.

Steps : (*i*) Take an assumed mean.

(*ii*) From the mid-point of each class deduct the assumed mean.

(*iii*) Multiply the respective frequencies of each class by these deviations and obtain the total Σfd.

(*iv*) Apply the formula : $\bar{X}=A+\frac{\Sigma fd}{N}$.

Calculate arithmetic mean by the short-cut method from the data of Illustration 4 (*a*).

CALCULATION OF ARITHMETIC MEAN

Marks	*Mid-point* m	*No. of Students* f	$(m-35)$ d	fd
0—10	5	5	−30	−150
10—20	15	10	−20	−200
20—30	25	25	−10	−250
30—40	35	30	0	0
40—50	45	20	+10	+200
50—60	55	10	+20	+200
		$N=100$		$\Sigma fd=-200$

$$\bar{X}=A+\frac{\Sigma fd}{N}=35-\frac{200}{100}=35-2=33.$$

In order to simplify the calculations, we can divide the deviations by class intervals, *i.e.*, calculate $(m-A)/i$ and then multiply by i in the formula for getting mean. The formula becomes :

$$\bar{X}=A+\frac{\Sigma fd}{N}\times i.$$

It may be pointed out that when class intervals are unequal we can simplify calculations by taking a common factor. In such a case we should use $(m-A)/C$ instead $(m-A)/i$ while making calculations.

From the data of Illustration 4 (*a*) compute arithmetic mean by step deviation method.

Solution. CALCULATION OF ARITHMETIC MEAN

Marks	*Mid-point* m	*No. of Students* f	$(m-35)$ d	$(m-35)/10$ d	fd
0—10	5	5	−30	−3	−15
10—20	15	10	−20	−2	−20
20—30	25	25	−10	−1	−25
30—40	35	30	0	0	0
40—50	45	[illegible]	+10	+1	+20
50—60	55	10	+20	+2	+20
		$N=100$			$\Sigma fd=-20$

$$\bar{X}=A+\frac{\Sigma fd}{N}\times i=35-\frac{20}{100}\times 10=35-2=33.$$

It is clear from above that all the three methods of finding arithmetic mean in continuous series give us the same answer. The direct method, though the simplest, involves more calculations when mid-points and frequencies are very large in magnitude. For example, observe the following data :

Income in Rs.	*No. of Persons*	*Income in Rs.*	*No. of Persons*
400—500	368	700—800	567
500—600	472	800—900	304
600—700	969		

In this case step deviation method would be far simpler. In fact, step deviation method should be adopted wherever possible because it minimises the calculations.

While computing mean in continuous series the mid-points of the various classes are taken as representative of that particular class. The reason is that when the data are grouped, the exact frequency with which each value of the variable occurs in the distribution is unknown. We only know the limits within which a certain number of frequencies occur. For example, when we say that the number of persons within the income group 400—500 is 50 we cannot say as to how many persons out of 50 are getting 401, 402, 403, etc. We, therefore, make an assumption while calculating arithmetic mean that the frequencies within each class are spread evenly over the range of the class interval, *i.e.*, there will be as many items below the mid-point as above it. Unless such an assumption is made the value of mean cannot be computed.

class is zero and the upper limit of the last class 150. In other words first class is 0—10 and the last one 100—150.

If the class intervals are of varying width, an effort should not be made to determine the lower limit of the lowest class and upper limit of the highest class. The use of median or mode would be better in such a case. Because of the difficulty of ascertaining lower limit and upper limit in open-end distributions it is suggested that in such distributions arithmetic mean should not be used.

Mathematical Properties of Arithmetic Mean

The following are a few important mathematical properties of the arithmetic mean :

1. The sum of the deviations of the items from the arithmetic mean (taking signs into account) is always zero, *i.e.*, $\Sigma(X-\bar{X})=0$*. This would be clear from the following example :

X	$(X-\bar{X})$
10	−20
20	−10
30	0
40	+10
50	+20
$\Sigma X=150$	$\Sigma(X-\bar{X})=0$

Here $\bar{X}=\frac{\Sigma X}{N}=\frac{150}{5}=30$. When the sum of the deviations from the actual mean, *i.e.*, 30, is taken it comes out to be zero. It is because of this property that the mean is characterised as a *point of balance*, *i.e.*, the sum of the positive deviations from it is equal to the sum of the negative deviations from it.

2. The sum of the squared deviations of the items from arithmetic mean is minimum, that is, less than the sum of the squared deviations of the items from any other value. The following example would verify the point :

X	$(X-\bar{X})$ $\bar{X}=4$	$(X-4)^2$
2	−2	4
3	−1	1
4	0	0
5	+1	1
6	+2	4
$\Sigma X=20$	$\Sigma(X-\bar{X})=0$	$\Sigma(X-\bar{X})^2=10$

* Algebraically the property $\Sigma(X-\bar{X})=0$ is derived from the fact that

$$N\bar{X}=\Sigma X.$$

Proof. $\Sigma(X-\bar{X})=(X_1-\bar{X})+(X_2-\bar{X})+\ldots+X_n=\bar{X}$

$$=(X_1+X_2.X_n-\bar{X}-\bar{X}\ldots n \text{ (times)}$$

$$=\Sigma X-\Sigma\bar{X}=N\bar{X}-N\bar{X}=0.$$

Also for a frequency distribution $\Sigma f(X-\bar{X})=0$.

The sum of the squared deviations is equal to 10 in the above case. If the deviations are taken from any other value the sum of the squared deviations would be greater than 10. For example, let us calculate the squares of the deviations of item from a value less than the arithmetic mean, say, 3.

X	$(X-3)$	$(X-3)^2$
2	−1	1
3	0	0
4	+1	1
5	+2	4
6	+3	9
		$\Sigma(X-3)^2=15$

It is clear that $(\Sigma X-3)^2$ is greater. This property that the sum of the squares of items is least from the mean is of immense use in regression analysis which shall be discussed later.

3. Since $\bar{X}=\frac{\Sigma X}{N}$, $N\bar{X}=\Sigma X$.

In other words, if we replace each item in the series by the mean, then the sum of these substitutions will be equal to the sum of the individual items. For example, in the discussion of first property $\Sigma X=150$ and the arithmetic mean 30. If for each item we substitute 30, we get the same total, *i.e.*. $30+30+30+30+30=150$.

This property is of great practical value. For example, if we know the average wage in a factory, say, Rs. 1,060 and the number of workers employed, say, 200, we can compute total wages bill from the relation $N\bar{X}=\Sigma X$. The total wage bill in this case would be $200\times 1{,}060$, *i.e.*, Rs. 2,12,000 which is equal to ΣX.

4. If we have the arithmetic mean and number of items of two or more than two related groups, we can compute combined average of these groups by applying the following formula :

$$\bar{X}_{12}=\frac{N_1\bar{X}_1+N_2\bar{X}_2}{N_1+N_2}$$

$\bar{X}_{12}$=combined mean of the two groups

$\bar{X}_1$=arithmetic mean of first group

$\bar{X}_2$=arithmetic mean of second group

N_1=number of items in the first group

N_2=number of items in the second group.

The following example shall illustrate the application of the above formula :

Illustration 6. The mean height of 25 male workers in a factory is 61 cms, and the mean height of 35 female workers in the same factory is 58 cms. Find the combined mean height of 60 workers in the factory. *(M.A. Econ., Punjab Univ., 1985)*

Solution. $\bar{X}_{12}=\dfrac{N_1\bar{X}_1+N_2\bar{X}_2}{N_1+N_2}$

$N_1=25,\ \bar{X}_1=61,\ N_2=35,\ \bar{X}_2=58$

$$\bar{X}_{12}=\frac{(25\times61)+(35\times58)}{25+35}=\frac{1525+2030}{60}=\frac{3555}{60}=59{\cdot}25.$$

Thus the combined mean height of 60 workers is 59·25 cm.

If we have to find out the combined mean of three sub-groups the above formula can be extended as follows :

$$\bar{X}_{123}=\frac{N_1\bar{X}_1+N_2\bar{X}_2+N_3\bar{X}_3}{N_1+N_2+N_3}$$

Merits and Limitations of Arithmetic Mean

Merits. Arithmetic mean is most widely used in practice because of the following reasons :

1. It is the simplest average to understand and easiest to compute. Neither the arraying of data as required for calculating median nor grouping of data as required for calculating mode is needed while calculating mean.

2. It is affected by the value of every item in the series.

3. It is defined by a rigid mathematical formula with the result that everyone who computes the average gets the same answer.

4. Being determined by a rigid formula, it lends itself to subsequent algebraic treatment better than the median or mode.

5. It is relatively reliable in the sense that it does not vary too much when repeated samples are taken from one and the same population, at least not as much as some other kind of statistical descriptions.

6. The mean is typical in the sense that it is the centre of gravity, balancing the values on either side of it.

7. It is a calculated value, and not based on position in the series.

Limitations. 1. Since the value of mean depends upon each and every item of the series, extreme items, *i.e.*, very small and very large items, unduly affect the value of the average. For example, if in a tutorial group there are 4 students and their marks in a test are 60, 70, 10 and 80 the average marks would be $\dfrac{60+70+10+80}{4}=\dfrac{220}{4}=55$. One single item, *i.e.*, 10, has reduced the average marks considerably. The smaller the number of observations, the greater is likely to be the impact of extreme values.

2. In a distribution with open-end classes the value of mean tcannot be computed without making assumptions regarding the size of lhe class interval of the open-end classes. If such classes contain a arge proportion of the values, then mean may be subject to substantial error. However, the values of the median and mode can be computed where there are open-end classes without making any assumptions about size of class interval.