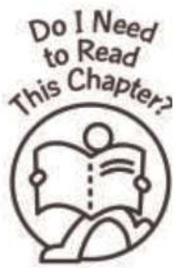
CHAPTER 03

Measures of Central Tendency

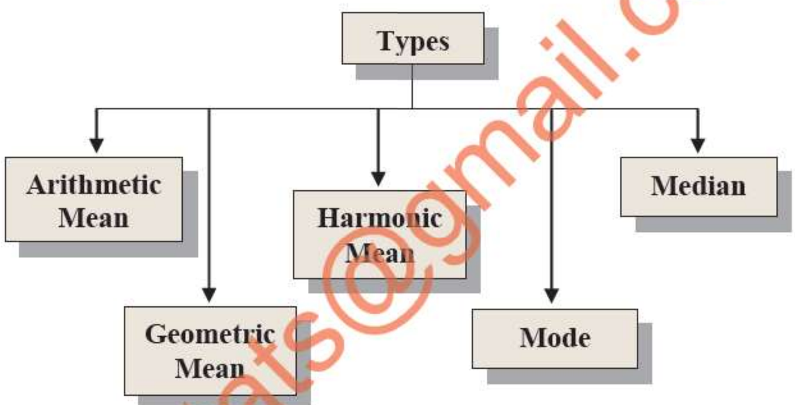
Chapter Contents



You should read this chapter if you need to learn about:

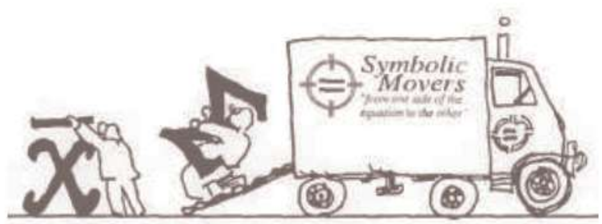
- Usually when two or more different data sets are to be compared it is necessary to condense the data, but for comparison the condensation of data set into a frequency distribution and visual presentation are not enough. It is then necessary to summarize the data set in a single value. Such a value usually somewhere in the center and represent the entire data set and hence it is called measure of central tendency or averages. Since a measure of central tendency (i.e. an average) indicates the location or the general position of the distribution on the X-axis therefore it is also known as a measure of location or position.

Types of Measure of Central Tendency



Arithmetic Mean or Simply Mean

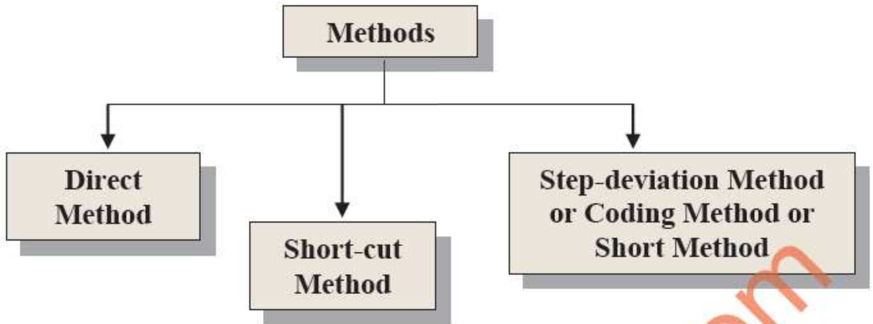
"A value obtained by dividing the sum of all the observations by the number of observations is called arithmetic mean"



$$Mean = \frac{Sum\ of\ All\ the\ Observations}{Number\ of\ Observations}$$

The mean is that central point where the sum of the negative deviations (absolute value) from the mean and the sum of the positive deviations from the mean are equal. This is why the mean is considered a measure of central tendency.

Methods of Finding Arithmetic Mean



Methods	Ungrouped data	Grouped data
Direct Method	$\bar{x} = \frac{\sum x_i}{n}$	$\bar{x} = \frac{\sum fx}{n}$; *Here* $n = \sum f$
Short cut Method	$\bar{x} = A + \frac{\sum D}{n}$	$\bar{x} = A + \frac{\sum fD}{n}$; *Here* $n = \sum f$
	Where $D = X_i - A$ and A is the provisional or assumed mean.	
Step deviation Method	$\bar{x} = A + \frac{\sum u}{n} \times h$	$\bar{x} = A + \frac{\sum fu}{n} \times h$; *Here* $n = \sum f$
	Where $u = \frac{X_i - A}{h}$ and h is the common width of the class intervals	

EXAMPLE 3.01

Find A.M from the following data: **(ungrouped data)**

2, 4, 6, 8, 10

Solution

Direct Method:

X
2
4
6
8
10
30

$$\bar{x} = \frac{\sum x_i}{n} = \frac{30}{5} = 6.0$$

REMEMBER

The Arithmetic mean is simply called Mean. We denoted Mean by $\bar{X}$ (read as "X Bar")

Short-cut Method:

X	$D = X_i - A$
2	-2
4	0
6	2
8	4
10	6
30	10

$$\bar{x} = A + \frac{\sum D}{n} = 4 + \frac{10}{5} = 6.0$$

(Let $A = 4$)

TIP

To compute the mean, round-off it one more decimal place than the original data values. For example, if the data are given in whole numbers, then the mean should be rounded-off to nearest tenth. If the data are given in tenths then the mean should be rounded-off to nearest hundredth and so on.

Step-deviation Method:

X	$u = \frac{X_i - A}{h}$
2	-3
4	-2
6	-1
8	0
10	1
30	-5

$$\bar{x} = A + \frac{\sum u}{n} \times h = 8 + \frac{(-5)}{5} \times 2 = 6.0$$

(Here $h = 2$ and let $A = 8$)

EXAMPLE 3.02

Find A.M from the following data: **(Discrete Grouped data)**

X	10	15	20	25	30
f	1	2	3	2	1

Solution

Direct Method:

X	f	fX
10	1	10
15	2	30
20	3	60
25	2	50
30	1	30
Total	9	180

$$\bar{x} = \frac{\sum fx}{n} = \frac{180}{9} = 20.0$$

(Here $n = \sum f = 9$)

TIP

In grouped data the number of observations "n" is equal to $\sum f$

Short-cut Method:

X	f	$D = X_i - A$	fD
10	1	-10	-10
15	2	-5	-10
20	3	0	0
25	2	5	10
30	1	10	10
Total	9	--	0

$$\bar{x} = A + \frac{\sum fD}{n} = 20 + \frac{0}{9} = 20.0$$

(Here $A = 20$ and $n = \sum f = 9$)

Step-deviation Method:

X	f	$u = \frac{X_i - A}{h}$	fu
10	1	-2	-2
15	2	-1	-2
20	3	0	0
25	2	1	2
30	1	2	2
Total	9	--	0

$$\bar{x} = A + \frac{\sum fu}{n} \times h = 20 + \frac{0}{9} \times 5 = 20.0$$

(Here $A = 20$, $h = 5$ and $n = \sum f = 9$)

EXAMPLE 3.03

Find A.M from the following data: **(Continuous Grouped data)**

Weight	11- 20	21- 30	31- 40	41-50	51-60
f	1	2	3	2	1

Solution

Direct Method:

Weight	f	*X (mid points)*	fX
11- 20	1	15.5	15.5
21- 30	2	25.5	51.0
31- 40	3	35.5	106.5
41-50	2	45.5	91.0
51-60	1	55.5	55.5
Total	9	--	319.5

$$\bar{x} = \frac{\sum fx}{n} = \frac{319.50}{9} = 35.50 \qquad (\text{here } n = \sum f = 9)$$

Short-cut Method:

Weight	*f*	*X (mid points)*	$D = X_i - A$	*fD*
11- 20	1	15.5	-20	-20
21- 30	2	25.5	-10	-20
31- 40	3	35.5	0	0
41-50	2	45.5	10	20
51-60	1	55.5	20	20
Total	9	--	--	0

$$\bar{x} = A + \frac{\sum fD}{n} = 35.5 + \frac{0}{9} = 35.50$$

(Here $A = 35.5$ and $n = \sum f = 9$)

Step-deviation Method:

Weight	*f*	*X (mid points)*	$u = \frac{X_i - A}{h}$	*fu*
11- 20	1	15.5	-2	-2
21- 30	2	25.5	-1	-2
31- 40	3	35.5	0	0
41-50	2	45.5	1	2
51-60	1	55.5	2	2
Total	9	--	--	0

$$\bar{x} = A + \frac{\sum fu}{n} \times h = 35.5 + \frac{0}{9} \times 10 = 35.50$$

(Here $A = 20$, $h = 10$ and $n = \sum f = 9$)

Historical Note

The concept of Arithmetic Mean has been first used by Greek astronomers in the third century BC.

But In 1755, Thomas Simpson officially proposed the use of Arithmetic Mean.

Test Yourself

Find the A.M from the following data:

1) 1, 3, 5, 7, 9, 11, 13, 15

2)

X	20	25	30	35	40
f	2	4	9	3	1

3)

Weight	21- 30	31- 40	41- 50	51-60	61-70
f	1	3	5	4	2

TIP

To find Mean of the population use the following formula:

$$\mu = \frac{\sum x}{N}$$

μ (meu)

Properties of Arithmetic Mean

The following are the properties of arithmetic mean:

- The mean of a constant is same constant.

4, 4, 4, 4, 4

$$\bar{x} = \frac{\sum xi}{n} = \frac{4+4+4+4+4}{5} = \frac{20}{5} = 4.0$$

- The sum of deviations from mean is equal to zero. i.e. $\sum(Xi - \bar{X}) = 0$

2, 4, 6, 8, 10

X	$(Xi - \bar{X})$
2	-4
4	-2
6	0
8	2
10	4
30	$0 = \sum(Xi - \bar{X})$

$$\bar{x} = \frac{\sum x_i}{n} = \frac{30}{5} = 6.0$$

$$\Rightarrow \sum(Xi - \bar{X}) = 0$$

- The sum of squared deviations from the mean is smaller than the sum of squared deviations from any arbitrary value or provisional mean. i.e. $\sum(xi - \bar{x})^2 < \sum(xi - A)^2$

2, 4, 6, 8, 10

X	$(Xi - \bar{X})$	$(Xi - \bar{X})^2$	$(Xi - A)$	$(Xi - A)^2$
2	-4	16	-2	4
4	-2	4	0	0
6	0	0	2	4
8	2	4	4	16
10	4	16	6	36
30	--	$40=\sum(Xi - \bar{X})^2$	--	$60=\sum(Xi - A)^2$

$$\bar{x} = \frac{\sum x_i}{n} = \frac{30}{5} = 6.0$$

Let A = 4

$$\Rightarrow \sum(Xi - \bar{X})^2 < \sum(Xi - A)^2$$

- The arithmetic mean is affected by the change of origin and scale i.e. when a constant is added to or subtracted from each value of a variable or if each value of a variable is multiplied or divided by a constant, then arithmetic mean is affected by these changes.

Variable	Mean
X_i	$\bar{X}$
$X_i \pm a$	$\bar{X} \pm a$
$a X_i$	$a\bar{X}$
$\frac{X_i}{a}$	$\frac{\bar{X}}{a}$

$$\bar{x} = \frac{\sum x_i}{n} = \frac{30}{5} = 6.0$$

Let $Y_i = 2X + 3$ (a = 3, b = 2)

X	Y = 2X+3
2	7
4	11
6	15
8	19
10	23
30	75

Now $\bar{Y} = \frac{\sum Y_i}{n} = \frac{75}{5} = 15.0$

therefore $\bar{Y} = b\bar{X} + a$ = (2) (6) + 3 = 15.0

- If k-subgroups consists of $n_1, n_2, \ldots, n_k$ observations having their respective means as $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_k$ then the mean of all the data or combined mean is denoted by $\bar{\bar{x}}$ or $\bar{x}_c$ and is defined by:

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

For example, if three sections of a statistics class containing 28, 32, and 35 students averaged 83, 80 and 76 respectively, on the same final examination. Then the combined mean for all 3 sections is:

$$n_1 = 28;\ \bar{X}_1 = 83$$
$$n_2 = 32;\ \bar{X}_2 = 80$$
$$n_3 = 35;\ \bar{X}_3 = 76$$

$$\bar{x}_c = \frac{n_1\bar{x}_1 + n_2\bar{x}_2 + n_3\bar{x}_3}{n_1 + n_2 + n_3} = \frac{(28)(83) + (32)(80) + (35)(76)}{28 + 32 + 35} = 79.4$$

Change of Origin

If we **add** a constant to each value of a variable or **subtract** a constant from each value of a variable, then this is called as **change of origin**. The **arithmetic mean** is **affected by** these changes but the **standard deviation** (will be discussed in **Chapter 04**) is **independent** of these changes. For example:

$$Mean(x) = \frac{\sum x}{n} = \frac{30}{5} = 6$$

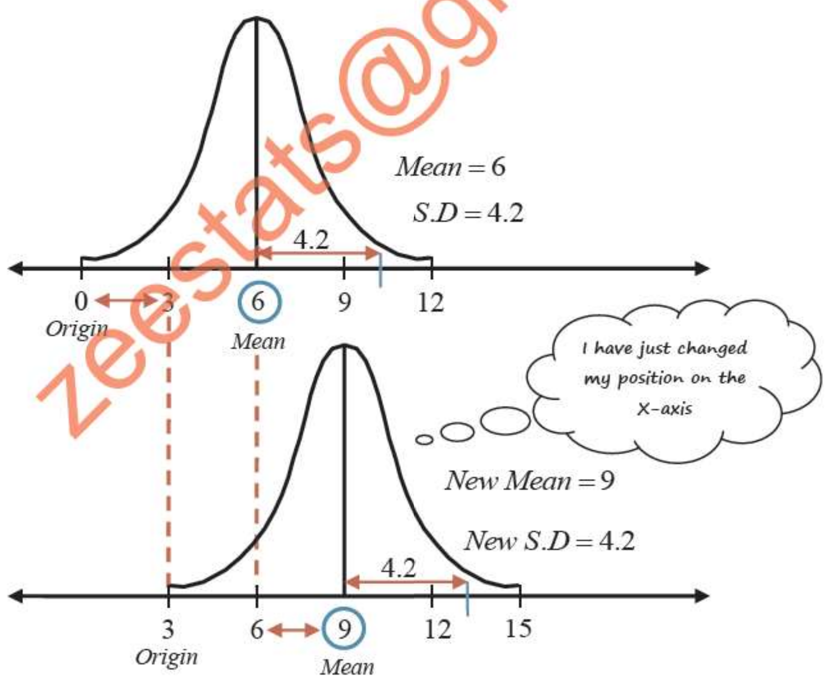
$$S.D(x) = \sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2} = \sqrt{\frac{270}{5} - \left(\frac{30}{5}\right)^2} = 4.2$$

$$Mean(y) = \frac{\sum y}{n} = \frac{45}{5} = 9$$

$$S.D(y) = \sqrt{\frac{\sum y^2}{n} - \left(\frac{\sum y}{n}\right)^2} = \sqrt{\frac{495}{5} - \left(\frac{45}{5}\right)^2} = 4.2$$

Old Variable		New Variable	
X	X^2	$Y = X+3$	Y^2
0	0	3	9
3	9	6	36
6	36	9	81
9	81	12	144
12	144	15	225
30	270	45	495

The following figure illustrates the idea of **change of origin**:



It is now clear, if we change the origin by adding "3" to each value of the variable, then the A.M will be affected by these changes but S.D will not be changed i.e.

$New\ Mean = (Old\ Mean + 3) = (6+3) = 9$ and $New\ S.D = Old\ S.D = 4.2$

Change of Scale

If each value of a variable is **multiply** or **divide** by a constant, then this is called as **change of scale**. The **arithmetic mean** and **standard deviation** are **affected by** these changes. For example:

$$Mean(x) = \frac{\sum x}{n} = \frac{30}{5} = 6$$

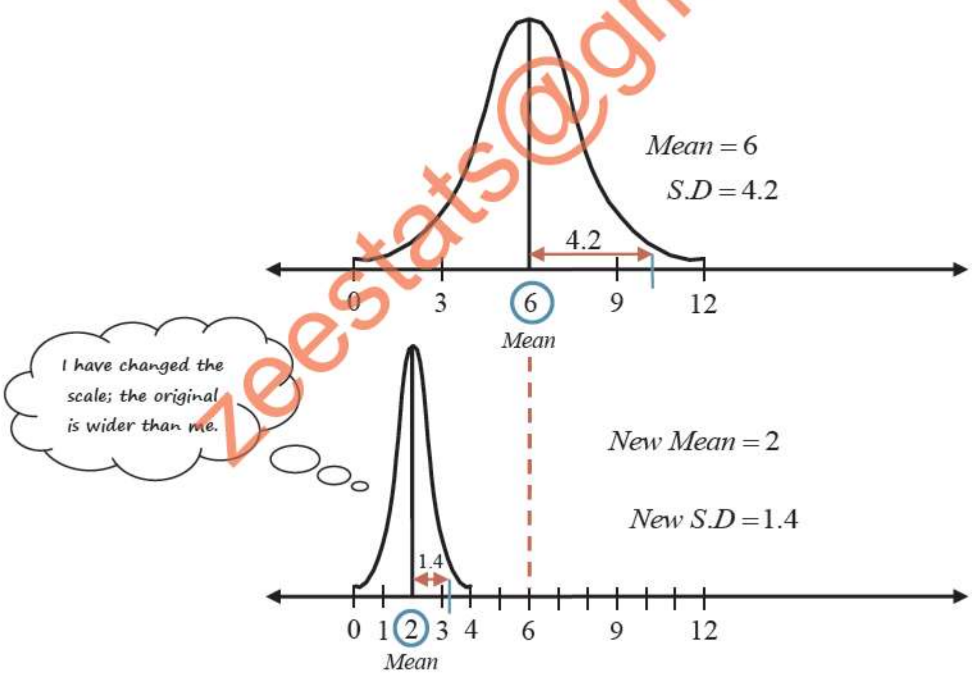
$$S.D(x) = \sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2} = \sqrt{\frac{270}{5} - \left(\frac{30}{5}\right)^2} = 4.2$$

$$Mean(y) = \frac{\sum y}{n} = \frac{10}{5} = 2$$

$$S.D(y) = \sqrt{\frac{\sum y^2}{n} - \left(\frac{\sum y}{n}\right)^2} = \sqrt{\frac{30}{5} - \left(\frac{10}{5}\right)^2} = 1.4$$

Old Variable		**New Variable**	
X	X^2	$Y = X/3$	Y^2
0	0	0	0
3	9	1	1
6	36	2	4
9	81	3	9
12	144	4	16
30	270	10	30

The following figure illustrates the idea of **change of scale**:



It is now clear, if we change the scale by dividing each value of the variable by "3" then both the A.M and S.D will be affected by these changes, such that:

$$New\ Mean = \frac{Old\ Mean}{3} = \frac{6}{3} = 2 \text{ and } New\ S.D = \frac{Old\ S.D}{3} = \frac{4.2}{3} = 1.4$$

Merits and Demerits of Arithmetic Mean

Merits

- The A.M is clearly defined by a mathematical formula.
- It is based on all the observations in the data and is easy to calculate.
- It is capable of further algebraic treatment.
- It is always unique, i.e. a set of data has only one mean.
- It is a relatively stable statistic with the fluctuations of sampling.
- It provide basis for statistical inference.

Demerits

- It is greatly affected by **extreme values** in the data.
- It cannot be calculated for qualitative data.
- If the grouped data have "open-end" classes, mean cannot be accurately computed.
- It is not an appropriate average for highly skewed distribution.

Weighted Arithmetic Mean

More Important

Less Important

Up till now we have discussed the simple A.M or in other words un-weighted A.M. In calculating arithmetic mean we assume that the values of a variable have equal importance. But it is not necessary that all the values have the same relative importance. Thus whenever it is required to find the mean of certain variables, which are not of equal importance, then we assign certain numerical quantities to these variables, which express their **relative importance**. Such numerical quantities are technically called the **weight**.

So it is obvious that we would modify the formula of the simple A.M and apply the formula of the weighted A.M i.e.

$$\bar{X}_w = \frac{\sum wx}{\sum w}$$

EXAMPLE 3.04

Calculate the weighted mean from the following data:

Item	*Expenditure (X)*	*Weights (W)*
Food	290	7.5
Rent	54	2.0
Clothing	98	1.5
Fuel & Light	75	1.0
Cosmetics	75	0.5

Solution

Since $\bar{X}_w = \dfrac{\sum wx}{\sum w}$

Item	Expenditure (x)	Weights (w)	wx
Food	290	7.5	2175.0
Rent	54	2.0	108.0
Clothing	98	1.5	147.0
Fuel & Light	75	1.0	75.0
Cosmetics	75	0.5	37.5
Total	--	12.5	2542.5

Therefore $\bar{X}_w = \dfrac{\sum wx}{\sum w} = \dfrac{2542.5}{12.5} = 203.4$

Test Yourself

Calculate the weighted mean from the following data:

Item	*Expenditure (X)*	*Weights (W)*
Food	390	9.5
Rent	44	3.0
Clothing	199	2.5
Fuel & Light	67	3.8
Other items	85	5.5

Geometric Mean

"The nth root of the product of "n" positive values is called geometric mean"

$$Geometric\ Mean = \sqrt[n]{Product\ of\ "n"\ Positive Values}$$

The following are the formulae of geometric mean:

Ungrouped data	Grouped data
$G = Antilog\left(\frac{\sum logx}{n}\right)$	$G = Antilog\left(\frac{\sum f\ logx}{n}\right)$;Here $n = \sum f$

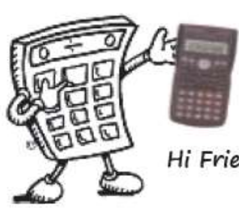
EXAMPLE 3.05

Find geometric mean from the following data: **(ungrouped data)**

5,8,10,12,15

Solution

X	$log\ X$
5	0.6990
8	0.9031
10	1.0000
12	1.0792
15	1.1761
Total	4.8573



$$G = Antilog\left(\frac{\sum logx}{n}\right)$$

$$= Antilog\left(\frac{4.8573}{5}\right) = 9.4$$

EXAMPLE 3.06

Find G.M from the following data: **(Discrete Grouped data)**

X	13	14	15	16	17
f	2	5	13	7	3

Solution

X	f	$\log X$	$f\log X$
13	2	1.1139	2.2279
14	5	1.1461	5.7306
15	13	1.1761	15.2892
16	7	1.2041	8.4288
17	3	1.2304	3.6913
Total	30	--	35.3679

$$G = Antilog\left(\frac{\sum flogx}{n}\right)$$

$$= Antilog\left(\frac{35.3679}{30}\right) = 15.1$$

EXAMPLE 3.07

Find G.M from the following data: **(Continuous Grouped data)**

Weights	65-84	85-104	105-124	125-144	145-164	165-184	185-204
f	9	10	17	10	5	4	5

Solution

Weights	f	X	$\log X$	$f\log X$
65-84	9	74.5	1.8722	16.8494
85-104	10	94.5	1.9754	19.7543
105-124	17	114.5	2.0588	34.9997
125-144	10	134.5	2.1287	21.2872
145-164	5	154.5	2.1889	10.9446
165-184	4	174.5	2.2418	8.9672
185-204	5	194.5	2.2889	11.4446
Total	60	--	--	124.2470

$$G = Antilog\left(\frac{\sum flogx}{n}\right)$$

$$= Antilog\left(\frac{124.2470}{60}\right) = 117.7$$

Test Yourself

Find the G.M from the following data:

1) 1, 3, 5, 7, 9, 11, 13, 15

2)

X	20	25	30	35	40
f	2	4	9	3	1

3)

Weight	21- 30	31- 40	41- 50	51-60	61-70
f	1	3	5	4	2

Merits and Demerits of Geometric Mean

BRAINTEASER

Merits

- The G.M is clearly defined by a mathematical formula.
- It is unique and based on all the observations.
- It is capable of further algebraic treatment.
- It is comparatively less affected by extreme values as compared to A.M.
- It gives equal weight to all the observations and is not much affected by fluctuations of sampling.

Demerits

- It is neither easy to calculate nor simple to understand.
- It vanishes if any observation is zero.
- It cannot be calculated for qualitative data.
- In case of negative values, it cannot be computed at all.
- If the grouped data have "open-end" classes, geometric mean cannot be accurately computed.

Harmonic Mean

"The reciprocal of the arithmetic mean of the reciprocals of the values is called harmonic mean"

$$Harmonic\ Mean = Reciprocal\ of\left(\frac{Sum\ of\ Reciprocal\ of\ the\ Values}{The\ Number\ of\ Values}\right)$$

The following are formulae of harmonic mean:

Ungrouped data	Grouped data
$H = \frac{n}{\sum\left(\frac{1}{x}\right)}$	$H = \frac{n}{\sum\left(\frac{f}{x}\right)}$; *Here* $n = \sum f$

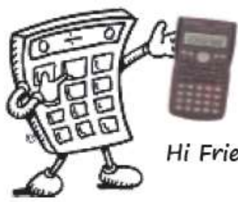
EXAMPLE 3.08

Find Harmonic mean from the following data: **(ungrouped data)**

5, 8, 10, 12, 15

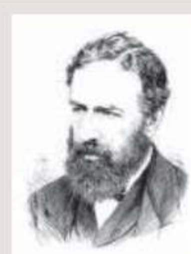
Solution

X	1/X
5	0.2000
8	0.1250
10	0.1000
12	0.0833
15	0.0667
Total	0.5750



$$H = \frac{n}{\sum\left(\frac{1}{x}\right)} = \frac{5}{0.5750} = 8.7$$

Historical Note



In 1874, Jevons William Stanley introduced the Geometric Mean and Harmonic Mean.

EXAMPLE 3.09

Find Harmonic mean from the following data: **(Discrete Grouped data)**

X	13	14	15	16	17
f	2	5	13	7	3

Solution

X	f	(f/X)
13	2	0.1538
14	5	0.3571
15	13	0.8667
16	7	0.4375
17	3	0.1765
Total	30	1.9916

$$H = \frac{n}{\sum\left(\frac{f}{x}\right)} = \frac{30}{1.9916} = 15.1$$

EXAMPLE 3.10

Find H.M from the following data: **(Continuous Grouped data)**

Weights	65-84	85-104	105-124	125-144	145-164	165-184	185-204
f	9	10	17	10	5	4	5

Solution

Weights	f	X	(f/X)
65-84	9	74.5	0.1208
85-104	10	94.5	0.1058
105-124	17	114.5	0.1485
125-144	10	134.5	0.0743
145-164	5	154.5	0.0324
165-184	4	174.5	0.0229
185-204	5	194.5	0.0257
Total	60	--	0.5304

$$H = \frac{n}{\sum\left(\frac{f}{x}\right)} = \frac{60}{0.5304} = 113.1$$

Test Yourself

Find the H.M from the following data:

1) 1, 3, 5, 7, 9, 11, 13, 15

2)

X	20	25	30	35	40
f	2	4	9	3	1

3)

Weight	21- 30	31- 40	41- 50	51-60	61-70
f	1	3	5	4	2

Merits and Demerits of Harmonic Mean

BRAINTEASER

Merits

- The H.M is clearly defined by a mathematical formula.
- It is unique and based on all the observations.
- It is capable of further algebraic treatment.
- It is comparatively less affected by extreme values as compared to A.M and G.M.
- It is not much affected by fluctuations of sampling.

Demerits

- It is neither easy to calculate nor simple to understand.
- It cannot be determined if any value is zero.
- It cannot be calculated for qualitative data.
- If the grouped data have "open-end" classes, geometric mean cannot be accurately computed.

REMEMBER

- If any value of the data is negative then G.M will become ill-defined and the remaining two averages relate each other inversely i.e. H.M > A.M.
- If any value of the data is zero, then H.M will become ill-defined and the G.M will be zero.
- A.M, G.M and H.M of two values "a" and "b" are:

$$A.M = \frac{a+b}{2}$$

$$G.M = (a \times b)^{1/2}$$

$$H.M = \frac{2ab}{a+b}$$

Relationship between Arithmetic Mean, Geometric Mean and Harmonic Mean

- A.M > G.M > H.M
- The three averages are exactly equal if the data set is constant i.e. A.M = G.M = H.M
- $(G.M)^2 \approx (A.M) \times (H.M)$

Consider the data:
2, 4, 6, 8, and 10

X	*log X*	*1/X*
2	0.3010	0.5000
4	0.6021	0.2500
6	0.7782	0.1667
8	0.9031	0.1250
10	1.0000	0.1000
30	3.5844	1.1417

$$\bar{x} = \frac{\sum x_i}{n} = \frac{30}{5} = 6$$

$$G = Antilog\left(\frac{\sum \log x}{n}\right)$$

$$= Antilog\left(\frac{3.5844}{5}\right) = 5.2$$

$$H = \frac{n}{\sum\left(\frac{1}{x}\right)} = \frac{5}{1.1417} = 4.4$$

Hence it is clear that: A.M > G.M > H.M

Consider the data:
10, 10, 10, 10, and 10

X	*log X*	*1/X*
10	1	0.1
10	1	0.1
10	1	0.1
10	1	0.1
10	1	0.1
50	5	0.5

$$\bar{x} = \frac{\sum x_i}{n} = \frac{50}{5} = 10$$

$$G = Antilog\left(\frac{\sum \log x}{n}\right)$$

$$= Antilog\left(\frac{5}{5}\right) = 10$$

$$H = \frac{n}{\sum\left(\frac{1}{x}\right)} = \frac{5}{0.5} = 10$$

Hence it is clear that: A.M = G.M = H.M

Historical Note

In 1970, the relationship between Arithmetic Mean, Geometric Mean and Harmonic Mean is described by Mitrinovic, D.S.

The A.M of two observations is 127.5 and their G.M is 60 find their H.M.

$$A.M = 127.5$$
$$G.M = 60$$
$$H.M = ?$$

$$(G.M)^2 \approx (A.M) \times (H.M) \Rightarrow H.M \approx \frac{(G.M)^2}{A.M} = \frac{(60)^2}{127.5} = 28.2$$

Mode

Mode in case of Ungrouped Data

"A value, that occurs most frequently in a data, is called mode"

e.g. 2, 3, 4, 2, 5, 6, 2, 7

Mode = 2

"If two or more values occur the same number of times but most frequently than the other values, then there is more than one mode"

e.g. 2, 9, 11, 9, 2, 13, 14, 7, 18

Mode = 2, 9

If each value occurs the same number of times, then there is no mode.

e.g.

1,2,3,4

(there is no mode)

5, 6, 5, 7, 6, 7

(there is no mode)

- *The data having one mode is called uni-modal distribution.*
- *The data having two modes is called bi-modal distribution.*
- *The data having more than two modes is called multi-modal distribution.*

Mode in case of Discrete Grouped Data

"A value, which has the largest frequency in a set of data, is called mode"

e.g.

X	41	42	43	44	45
f	1	3	5	2	1

Mode = 43
(Against the maximum frequency)

Mode in case of Continuous Grouped Data

In case of continuous grouped data, mode would lie in the class that carries the highest frequency. This class is called the modal class. The formula used to compute the value of mode, is given below:

$$Mode = l + \frac{f_m - f_1}{(f_m - f_1) + (f_m - f_2)} \times h$$

Where l = lower class boundary of the modal class
h = class-width of the modal class
f_m = frequency of the modal class
f_1 = frequency of the class preceding the modal class
f_2 = frequency of the class following the modal class

Historical Note

In 1894, Karl Pearson used the term "Mode"

EXAMPLE 3.11

Find mode from the following data:

Marks	30-39	40-49	50-59	60-69	70-79	80-89	90-99
No. of Students	8	87	190	304	211	85	20

Solution

Since $$Mode = l + \frac{f_m - f_1}{(f_m - f_1) + (f_m - f_2)} \times h$$

Marks	No. of Students	Class boundaries
30-39	8	29.5-39.5
40-49	87	39.5-49.5
50-59	190	49.5-59.5
60-69	304	59.5-69.5
70-79	211	69.5-79.5
80-89	85	79.5-89.5
90-99	20	89.5-99.5

Modal class: 59.5 — 69.5

$l = 59.5, f_1 = 190, f_2 = 211, f_m = 304,$

$h = 69.5 - 59.5 = 10$

$$Mode = l + \frac{f_m - f_1}{(f_m - f_1) + (f_m - f_2)} \times h$$

$$= 59.5 + \frac{304 - 190}{(304 - 190) + (304 - 211)} \times 10 = 65$$

Mode Graphically

- Construct a Histogram form the continuous grouped data.
- Locate the modal class i.e. the class with highest rectangle.
- Draw a line from top right hand corner of the modal class rectangle to the point where the top of the next adjacent rectangle to the left- touches. Similarly, join the top left hand corner of the modal class rectangle to the point where the top of the next adjacent rectangle to the right -touches.
- From the intersection of these two lines draw a perpendicular on X-axis.
- Mode is the point where the perpendicular meets the X-axis.

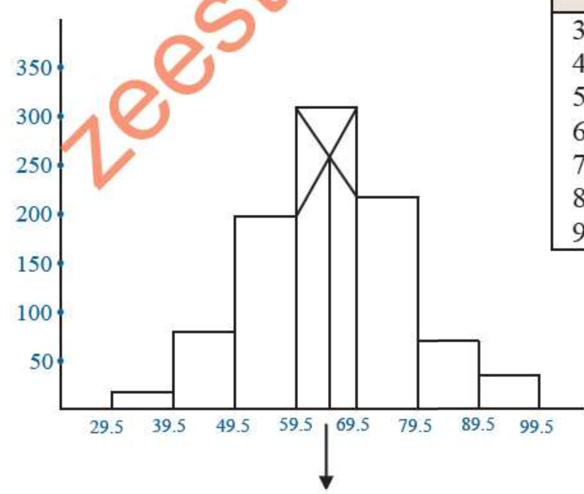
EXAMPLE 3.12

Find mode graphically from the following data:

Marks	30-39	40-49	50-59	60-69	70-79	80-89	90-99
No. of Students	8	87	190	304	211	85	20

Solution

Marks	No. of Students	Class boundaries
30-39	8	29.5-39.5
40-49	87	39.5-49.5
50-59	190	49.5-59.5
60-69	304	59.5-69.5
70-79	211	69.5-79.5
80-89	85	79.5-89.5
90-99	20	89.5-99.5



Mode = 65.0072

Merits and Demerits of Mode

Merits

- It is simple to understand and easy to calculate.
- In some cases it may be obtained by just inspection.
- It is not affected by extreme values.
- It is also useful for qualitative data.
- It can be located even in open-end classes.

Demerits

- It is not clearly defined by a mathematical formula.
- It may not exist in some cases.
- It is non-unique for all types of data.
- It is not capable of further algebraic treatment.
- It is not based on all the observations.
- It is unsatisfactory for statistical inference.

Test Yourself

Find the Mode from the following data:

1) 1, 3, 5, 7, 7, 11, 13, 7

2)

X	20	25	30	35	40
f	2	4	9	3	1

3)

Weight	21- 30	31- 40	41- 50	51-60	61-70
f	1	3	5	4	2

Median

"When the observations are arranged in ascending or descending order, then a value, that divides a distribution into two equal parts, is called median"

Median in case of Ungrouped Data	
In this case we first arrange the observations in **increasing** or **decreasing** order then we use the following formulae for Median:	
If "n" is odd	$Median = size\ of \left(\frac{n+1}{2}\right) th\ observation$
If "n" is even	$Median = \dfrac{size\ of \left\{\left(\frac{n}{2}\right) th + \left(\frac{n}{2}+1\right) th\right\}\ observation}{2}$

EXAMPLE 3.13

Find Median from the following data:

3,4,5,8,2,9,7,6,10

Solution

Ascending order: 2,3,4,5,6,7,8,9,10 (n = 9 odd)

$$Median = size\ of \left(\frac{n+1}{2}\right) th\ observation$$

$$= size\ of \left(\frac{9+1}{2}\right) th\ observation$$

$$= size\ of\ 5th\ observation = 6$$

The number of values above the median balances (equals) the number of values below the median i.e. 50% of the data falls above and below the median.

EXAMPLE 3.14

Find Median from the following data:

13,14,15,18,12,19,17,16,10,20

Solution

Ascending order: 10,12,13,14,15,16,17,18,19,20

(n = 10 even)

$$Median = \frac{size\ of \left\{\left(\frac{n}{2}\right)th + \left(\frac{n}{2}+1\right)th\right\} observation}{2}$$

$$= \frac{size\ of \left\{\left(\frac{10}{2}\right)th + \left(\frac{10}{2}+1\right)th\right\} observation}{2}$$

$$= \frac{size\ of\ [5th + 6th]\ observation}{2}$$

$$= \frac{15+16}{2} = 15.5$$

Median in case of Discrete Grouped Data

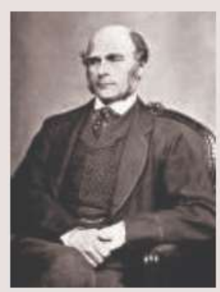
In case of discrete grouped data, first we find the cumulative frequencies and then use the following formula for Median:

$$Median = size\ of \left(\frac{n+1}{2}\right)th\ observation$$

Here $n = \sum f$

Historical Note

The concept of Median was used by Gauss at the beginning of 19th century.



Around 1874 Francis Galton first introduced Median as statistical concept

EXAMPLE 3.15

Find Median from the following data:

X	20	21	22	23	24	25
f	1	3	5	2	2	2

Solution

X	*f*	*Cumulative Frequency*
20	1	1
21	3	4
22 ←	5	9
23	2	11
24	2	13
25	2	15
Total	15	--

$$Median = size\ of \left(\frac{n+1}{2}\right) th\ observation$$

$$= size\ of \left(\frac{15+1}{2}\right) th\ observation$$

$$= size\ of\ 8th\ observation = 22$$

EXAMPLE 3.16

Find Median from the following data:

X	41	42	43	44	45	46
f	2	4	4	2	1	3

Solution

X	f	Cumulative Frequency
41	2	2
42	4	6
43 ←	4	10
44	2	12
45	1	13
46	3	16
Total	16	--

$$Median = size\ of \left(\frac{n+1}{2}\right) th\ observation$$

$$= size\ of \left(\frac{16+1}{2}\right) th\ observation$$

$$= size\ of\ 8.5th\ observation = 43$$

Median in case of continuous Grouped Data

In continuous grouped data, when we are finding median, we first construct the class boundaries if the classes are discontinuous. Then we find cumulative frequencies and then we use the following two steps:

- First we determine the median class using n/2.
- When the median class is determined, then the following formula is used to find the value of median. i.e.

$$Median = l + \frac{h}{f}\left(\frac{n}{2} - C\right); \quad Here\ n = \Sigma f$$

Where l = lower class boundary of the median class
h = class-width of the median class
f = frequency of the median class
C = cumulative frequency of the class preceding the median class.

EXAMPLE 3.17

Find Median from the following data:

Marks	30-39	40-49	50-59	60-69	70-79	80-89	90-99
No. of Students	8	87	190	304	211	85	20

Solution

Marks	No. of Students	C.F	Class boundaries
30-39	8	8	29.5-39.5
40-49	87	95	39.5-49.5
50-59	190	285	49.5-59.5
60-69	304	589 →	59.5-69.5
70-79	211	800	69.5-79.5
80-89	85	885	79.5-89.5
90-99	20	905	89.5-99.5
Total	905	--	--

Step 1:

$$Median = Size\ of\left(\frac{n}{2}\right)th\ observation$$

$$= Size\ of\left(\frac{905}{2}\right)th\ observation$$

$$= Size\ of\ 452.5th\ observation$$

And since 452.5^{th} observation lies in the class (59.5-69.5); hence this is the median class.

Here l = 59.5, f = 304, C = 285, h = 10

Step 2:

$$Median = l + \frac{h}{f}\left(\frac{n}{2} - C\right)$$

$$Median = 59.5 + \frac{10}{304}(452.5 - 285) = 65$$

Test Yourself

Find the Median from the following data:

1) 1, 3, 5, 7, 7, 11, 13, 7, 6

2) 30, 44, 34, 46, 55, 47, 20, 58

3)

X	20	25	30	35	40
f	2	4	9	3	1

4)

Weight	21- 30	31- 40	41- 50	51-60	61-70
f	1	3	5	4	2

Graphic Representation of Median

- Draw an ogive on the basis of "less than" or "more than" type.
- Compute (n/2) and locate this point on vertical scale (y-axis).
- Draw a perpendicular from the located point to the ogive.
- Now draw a perpendicular on x-axis from the point where the first perpendicular cuts the ogive.
- The point at which the perpendicular will intersect the x-axis will be the Median of the distribution.

EXAMPLE 3.18

Find Median graphically from the following data:

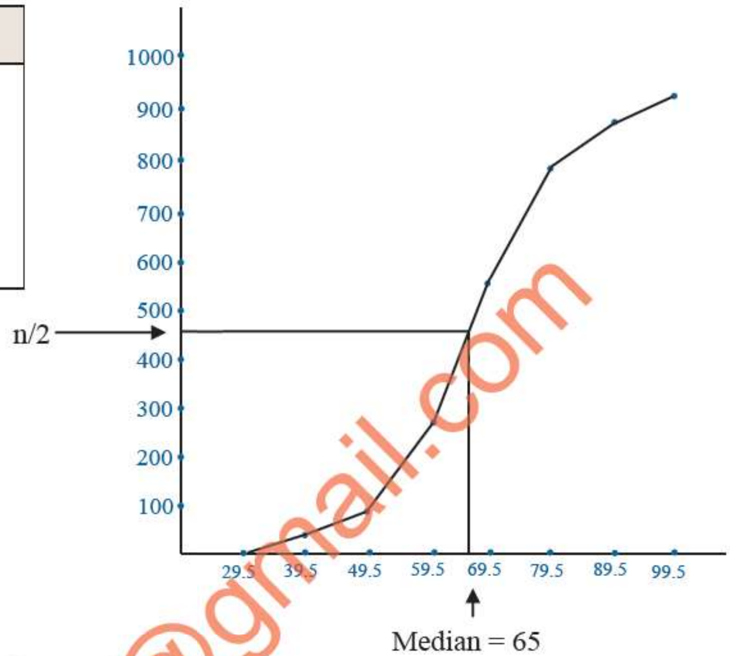
Marks	30-39	40-49	50-59	60-69	70-79	80-89	90-99
No. of Students	8	87	190	304	211	85	20

Solution

Marks	*No. of Students*	*C.F*	*Class boundaries*
30-39	8	8	29.5-39.5
40-49	87	95	39.5-49.5
50-59	190	285	49.5-59.5
60-69	304	589	59.5-69.5
70-79	211	800	69.5-79.5
80-89	85	885	79.5-89.5
90-99	20	905	89.5-99.5
Total	905	--	--

Here we construct "less than" cumulative frequency distribution:

Less than Class boundaries	*C.F*
Less than 29.5	0
Less than 39.5	8
Less than 49.5	95
Less than 59.5	285
Less than 69.5	589
Less than 79.5	800
Less than 89.5	885
Less than 99.5	905



Merits and Demerits of Median

Merits

- It is simple to understand and easy to calculate.
- It is not affected by extreme values.
- It is also useful for qualitative data.
- It can be located even in open-end classes.
- Like mean it always exists and is unique for any set of data.
- It is the most appropriate average in highly skewed distribution.

Demerits

- It is not clearly defined by a mathematical formula.
- It is not capable of further algebraic treatment.
- It is not based on all the observations.
- It is necessary to arrange the values in an array before finding the median, which is a tedious (boring) work.
- It is unsatisfactory for statistical inference.

WARNING!

- *It will be incorrect if we get the answer of an average out side the range of the data.*
- *Whenever you hear the word "average", be aware that the word may not always be referring to the mean. It may refer to Median and Mode etc.*

The averages that are obtained by using mathematical formulae are called mathematical averages e.g.

- Arithmetic Mean
- Harmonic Mean
- Geometric Mean

The averages that are obtained by simple inspection of the data are called positional averages e.g.

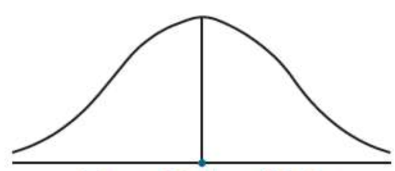
- Mode
- Median

All these averages are affected by the change of origin and scale.

Symmetrical Distribution

"A distribution is said to be symmetric if the values of mean, median and mode are equal" i.e.

$$\text{Mean} = \text{Median} = \text{Mode}$$



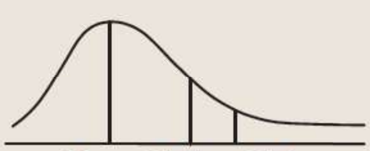
Mean = Median = Mode

In symmetrical distribution the sum of the deviation from the mean, mode or median is zero. The shape of such a distribution is always in the form of a bell, as shown in the figure.

"For symmetric distribution, we know that the values of mean, median and mode are equal, but if these values differ, then the distribution is said to be skewed or asymmetric"

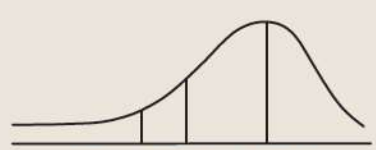
The following figures show the skewed distribution:

+ve Skewness



Mean > Median > Mode

–ve Skewness



Mean < Median < Mode

Empirical Relation between Mean, Median and Mode

"The difference between mean and mode is three times the difference between mean and median" i.e.

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

OR

"The difference between median and mode is twice the difference between mean and median".

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

If two averages are given then we can find the third one using:

- $Mean = \dfrac{3Median - Mode}{2}$
- $Median = \dfrac{Mode + 2Mean}{3}$
- $Mode = 3Median - 2Mean$

If Mean = 28.5 and Median = 30 then by Empirical Relation:

$Mode = 3Median - 2Mean \Rightarrow Mode = 3(30) - 2(28.5) = 33$

Quartiles

"When the observations are arranged in increasing order then the values, that divide the whole data into four (4) equal parts, are called quartiles"

These values are denoted by Q_1, Q_2 and Q_3.

It is to be noted that 25% of the data falls below Q_1, 50% of the data falls below Q_2 and 75% of the data falls below Q_3.

Quartiles, Deciles and Percentiles are also called Quantiles or Fractiles.

Deciles

"When the observations are arranged in increasing order then the values, that divide the whole data into ten (10) equal parts, are called deciles"

These values are denoted by $D_1, D_2, \ldots, D_9$.

It is to be noted that 10% of the data falls below D_1, 20% of the data falls below $D_2, \ldots,$ and 90% of the data falls below D_9.

Percentiles

"When the observations are arranged in increasing order then the values, that divide the whole data into hundred (100) equal parts, are called percentiles"

REMEMBER

For a data 2nd quartile, 5th decile and 50th percentile are equal to Median i.e.
$Q_2 = D_5 = P_{50} =$ Median

These values denoted by $P_1, P_2, \ldots, P_{99}$.

It is to be noted that 1% of the data falls below P_1, 2% of the data falls below $P_2, \ldots$, and 99% of the data falls below P_{99}.

Measures	Data Type	Formulas
Quartiles $j = 1, 2, 3$	**Ungrouped Data**	$Q_j = size\ of\left(\frac{j(n+1)}{4}\right)th\ observation$
	Discrete Grouped data	$Q_j = size\ of\left(\frac{j(n+1)}{4}\right)th\ observation$; Here $n = \sum f$
Deciles $j = 1, 2, \ldots, 9$	**Ungrouped Data**	$D_j = size\ of\left(\frac{j(n+1)}{10}\right)th\ observation$
	Discrete Grouped data	$D_j = size\ of\left(\frac{j(n+1)}{10}\right)th\ observation$; Here $n = \sum f$
Percentiles $j = 1, 2, \ldots, 99$	**Ungrouped Data**	$P_j = size\ of\left(\frac{j(n+1)}{100}\right)th\ observation$
	Discrete Grouped data	$P_j = size\ of\left(\frac{j(n+1)}{100}\right)th\ observation$; Here $n = \sum f$

Continuous Grouped Data	
In continuous grouped data, we use the following two steps:	
Quartiles	• First we determine the j^{th} quartile class using jn/4. • When the j^{th} quartile class is determined, then the following formula is used to find the value of j^{th} quartile i.e. $Q_j = l + \frac{h}{f}\left(\frac{jn}{4} - C\right)$; Here $n = \Sigma f$ l = lower class boundary of the j^{th} quartile class h = class-width of the j^{th} quartile class f = frequency of the j^{th} quartile class C = cumulative frequency of the class preceding the j^{th} quartile class.
Deciles	• First we determine the j^{th} decile class using jn/10. • When the j^{th} decile class is determined, then the following formula is used to find the value of the j^{th} decile. i.e. $D_j = l + \frac{h}{f}\left(\frac{jn}{10} - C\right)$; Here $n = \Sigma f$ l = lower class boundary of the j^{th} decile class h = class-width of the j^{th} decile class f = frequency of the j^{th} decile class C = cumulative frequency of the class preceding the j^{th} decile class.
Percentiles	• First we determine j^{th} percentile class using jn/100. • When the j^{th} percentile class is determined, then the following formula is used to find the value of the j^{th} percentile. i.e. $P_j = l + \frac{h}{f}\left(\frac{jn}{100} - C\right)$; Here $n = \Sigma f$ l = lower class boundary of the j^{th} percentile class h = class-width of the j^{th} percentile class f = frequency of the j^{th} percentile class C = cumulative frequency of the class preceding the j^{th} percentile class.

EXAMPLE 3.19

Find Q_1, Q_3, D_5, and P_{50} from the following data:

50,51,52,53,54,55,56,57,58,59,60; (n = 11)

Solution

$$Q_1 = \text{size of}\left(\frac{1(n+1)}{4}\right)\text{th observation}$$
$$\Rightarrow Q_1 = \text{size of}\left(\frac{1(11+1)}{4}\right)\text{th observation}$$
$$= \text{size of 3th observation} = 52$$

$$Q_3 = \text{size of}\left(\frac{3(n+1)}{4}\right)\text{th observation}$$
$$\Rightarrow Q_3 = \text{size of}\left(\frac{3(11+1)}{4}\right)\text{th observation}$$
$$= \text{size of 9th observation} = 58$$

$$D_5 = \text{size of}\left(\frac{5(n+1)}{10}\right)\text{th observation}$$
$$\Rightarrow D_5 = \text{size of}\left(\frac{5(11+1)}{10}\right)\text{th observation}$$
$$= \text{size of 6th observation} = 55$$

$$P_{50} = \text{size of}\left(\frac{50(n+1)}{100}\right)\text{th observation}$$
$$\Rightarrow P_{50} = \text{size of}\left(\frac{50(11+1)}{100}\right)\text{th observation}$$
$$= \text{size of 6th observation} = 55$$

EXAMPLE 3.20

Find Q_1, Q_3, D_6 and P_{80} from the following data:

150,151,152,153,154,155,156,157,158,159 (n = 10)

Solution

$$Q_1 = \text{size of}\left(\frac{1(n+1)}{4}\right)\text{th observation}$$
$$\Rightarrow Q_1 = \text{size of}\left(\frac{1(10+1)}{4}\right)\text{th observation}$$
$$= \text{size of 2.75th observation}$$
$$= \text{size of}\left[\text{2nd} + 0.75(\text{3rd - 2nd})\right]\text{observation}$$
$$= 151 + 0.75(152 - 151) = 151.75$$

$$Q_3 = \text{size of}\left(\frac{3(n+1)}{4}\right)\text{th observation}$$

$$\Rightarrow Q_3 = \text{size of}\left(\frac{3(10+1)}{4}\right)\text{th observation}$$

$$= \text{size of } 8.25\text{th observation}$$

$$= \text{size of}\left[8th + 0.25(9th - 8th)\right]\text{observation}$$

$$= 157 + 0.25(158 - 157) = 157.25$$

$$D_6 = \text{size of}\left(\frac{6(n+1)}{10}\right)\text{th observation}$$

$$\Rightarrow D_6 = \text{size of}\left(\frac{6(10+1)}{10}\right)\text{th observation}$$

$$= \text{size of } 6.6\text{th observation}$$

$$= \text{size of}\left[6th + 0.6(7th - 6th)\right]\text{observation}$$

$$= 155 + 0.6(156 - 155) = 155.6$$

$$P_{80} = \text{size of}\left(\frac{80(n+1)}{100}\right)\text{th observation}$$

$$\Rightarrow P_{80} = \text{size of}\left(\frac{80(10+1)}{100}\right)\text{th observation}$$

$$= \text{size of } 8.8\text{th observation}$$

$$= \text{size of}\left[8th + 0.8(9th - 8th)\right]\text{observation}$$

$$= 157 + 0.8(158 - 157) = 157.8$$

EXAMPLE 3.21

Find Q_1, Q_3, D_4 and P_{60} from the following data:

X	20	21	22	23	24	25
f	1	3	5	2	2	2

Solution

X	f	Cumulative Frequency
20	1	1
21 ←	3	4
22 ←	5	9
23 ←	2	11
24 ←	2	13
25	2	15
Total	15	--

$$Q_1 = size\ of\left(\frac{1(n+1)}{4}\right)th\ observation$$
$$\Rightarrow Q_1 = size\ of\left(\frac{1(15+1)}{4}\right)th\ observation$$
$$= size\ of\ 4th\ observation = 21$$

$$Q_3 = size\ of\left(\frac{3(n+1)}{4}\right)th\ observation$$
$$\Rightarrow Q_3 = size\ of\left(\frac{3(15+1)}{4}\right)th\ observation$$
$$= size\ of\ 12th\ observation = 24$$

$$D_4 = size\ of\left(\frac{4(n+1)}{10}\right)th\ observation$$
$$\Rightarrow D_4 = size\ of\left(\frac{4(15+1)}{10}\right)th\ observation$$
$$= size\ of\ 6.4th\ observation = 22$$

$$P_{60} = size\ of\left(\frac{60(n+1)}{100}\right)th\ observation$$
$$\Rightarrow P_{60} = size\ of\left(\frac{60(15+1)}{100}\right)th\ observation$$
$$= size\ of\ 9.6th\ observation = 23$$

EXAMPLE 3.22

Find Q_1, Q_3, D_8 and P_{40} from the following data:

Marks	30-39	40-49	50-59	60-69	70-79	80-89	90-99
No. of Students	8	87	190	304	211	85	20

Solution

Marks	No. of Students	C.F	Class boundaries
30-39	8	8	29.5-39.5
40-49	87	95	39.5-49.5
50-59	190	285 →	49.5-59.5
60-69	304	589 →	59.5-69.5
70-79	211	800 →	69.5-79.5
80-89	85	885	79.5-89.5
90-99	20	905	89.5-99.5
Total	905	--	--

Step 1

$$Q_1 = Size\ of\left(\frac{1\times n}{4}\right)th\ observation$$

$$\Rightarrow \quad Q_1 = Size\ of\left(\frac{1\times 905}{4}\right)th\ observation$$

$$= Size\ of\left(\frac{905}{4}\right)th\ observation$$

$$= Size\ of\ 226.25th\ observation$$

And since 226.25th observation lies in the class (49.5-59.5); hence this is the lower quartile class.

Here $l = 49.5, f = 190, C = 95, h = 10$

Step 2

Now using the following formula:

$$Q_1 = l + \frac{h}{f}\left(\frac{1\times n}{4} - C\right)$$

$$Q_1 = 49.5 + \frac{10}{190}\left(\frac{1\times 905}{4} - 95\right)$$

$$= 49.5 + \frac{10}{190}(226.25 - 95)$$

$$= 56.40$$

Step 1

$$Q_3 = Size\ of\left(\frac{3\times n}{4}\right)th\ observation$$

$$\Rightarrow \quad Q_3 = Size\ of\left(\frac{3\times 905}{4}\right)th\ observation$$

$$= Size\ of\left(\frac{2715}{4}\right)th\ observation$$

$$= Size\ of\ 678.75th\ observation$$

And since 678.75th observation lies in the class (69.5-79.5); hence this is the upper quartile class.

Here $l = 69.5, f = 211, C = 589, h = 10$

Step 2

Now using the following formula:

$$Q_3 = l + \frac{h}{f}\left(\frac{3n}{4} - C\right)$$

$$Q_3 = 69.5 + \frac{10}{211}\left(\frac{3\times 905}{4} - 589\right)$$

$$= 69.5 + \frac{10}{211}(678.75 - 589)$$

$$= 73.75$$

Step 1

$$D_8 = Size\ of\left(\frac{8n}{10}\right)th\ observation$$

$$\Rightarrow \quad D_8 = Size\ of\left(\frac{8\times 905}{10}\right)th\ observation$$

$$= Size\ of\left(\frac{7240}{10}\right)th\ observation$$

$$= Size\ of\ 724th\ observation$$

And since 724th observation lies in the class (69.5-79.5); hence this is the 8th decile class.

Here $l = 69.5, f = 211, C = 589, h = 10$

Step 2

Now using the following formula:

$$D_8 = l + \frac{h}{f}\left(\frac{8n}{10} - C\right)$$

$$D_8 = 69.5 + \frac{10}{211}\left(\frac{8\times 905}{10} - 589\right)$$

$$= 69.5 + \frac{10}{211}(724 - 589)$$

$$= 75.89$$

Step 1

$$P_{40} = Size\ of\left(\frac{40n}{100}\right)th\ observation$$

$$\Rightarrow \quad P_{40} = Size\ of\left(\frac{40\times 905}{100}\right)th\ observation$$

$$= Size\ of\left(\frac{36200}{100}\right)th\ observation$$

$$= Size\ of\ 362th\ observation$$

And since 362th observation lies in the class (59.5-69.5); hence this is the 40th percentile class.

Here $l = 59.5, f = 304,\ C = 285,\ h = 10$

Step 2

Now using the following formula:

$$P_{40} = l + \frac{h}{f}\left(\frac{40n}{100} - C\right)$$

$$P_{40} = 59.5 + \frac{10}{304}\left(\frac{40\times 905}{100} - 285\right)$$

$$= 59.5 + \frac{10}{304}(362 - 285)$$

$$= 62.03$$

Main Objects of Average

- The main object (purpose) of the average is to give a bird's eye view (summary) of the statistical data. The average removes all the unnecessary details of the data and gives a concise (to the point or short) picture of the huge data under investigation.
- Average is also of great use for the purpose of comparison (i.e. the comparison of two or more groups in which the units of the variables are same) and for the further analysis of the data.
- Averages are very useful for computing various other statistical measures such as dispersion, skewness, kurtosis etc.

Requisites (desirable qualities) of a Good Average

An average will be considered as good if:

- It is mathematically defined.
- It utilizes all the values given in the data.
- It is not much affected by the extreme values.
- It can be calculated in almost all cases.
- It can be used in further statistical analysis of the data.
- It should avoid to give misleading results.

Uses of Averages in Different Situations

- **A.M** is an appropriate average for **all the situations** where there are no extreme values in the data.
- **G.M** is an appropriate average for calculating **average percent increase** in sales, population, production, etc. It is one of the best averages for the construction of **index numbers**.
- **H.M** is an appropriate average for calculating the **average rate of increase** of profits of a firm or finding average speed of a journey or the average price at which articles are sold.
- **Mode** is an appropriate average in case of **qualitative data** e.g. the opinion of an average person; he is probably referring to the most frequently expressed opinion which is the modal opinion.
- **Median** is an appropriate average in a **highly skewed distribution** e.g. in the distribution of wages, incomes etc.

Prove that: $\sum(xi-\overline{x})^2 < \sum(xi-A)^2$

Proof: *Taking* $\sum(xi-A)^2 = \sum(xi-\overline{x}+\overline{x}-A)^2$

$$= \sum[(xi-\overline{x})+(\overline{x}-A)]^2$$

$$= \sum[(xi-\overline{x})^2+(\overline{x}-A)^2+2(xi-\overline{x})(\overline{x}-A)]$$

$$= \sum(xi-\overline{x})^2+\sum(\overline{x}-A)^2+2\sum(xi-\overline{x})(\overline{x}-A)$$

$$= \sum(xi-\overline{x})^2+n(\overline{x}-A)^2+2(\overline{x}-A)\sum(xi-\overline{x})$$

$$= \sum(xi-\overline{x})^2+n(\overline{x}-A)^2 \qquad \left[\because \sum(xi-\overline{x})=0\right]$$

$$\Rightarrow \sum(xi-A)^2 < \sum(xi-\overline{x})^2 \qquad \left[\because n(\overline{x}-A)^2>0\right]$$