# 6. MEASURES OF CENTRAL TENDENCY

**Measures of Central Tendency:**

In the study of a population with respect to one in which we are interested we may get a large number of observations. It is not possible to grasp any idea about the characteristic when we look at all the observations. So it is better to get one number for one group. That number must be a good representative one for all the observations to give a clear picture of that characteristic. Such representative number can be a central value for all these observations. This central value is called a measure of central tendency or an average or a measure of locations. There are five averages. Among them mean, median and mode are called simple averages and the other two averages geometric mean and harmonic mean are called special averages.

The meaning of average is nicely given in the following definitions.

"A measure of central tendency is a typical value around which other figures congregate."

"An average stands for the whole group of which it forms a part yet represents the whole."

"One of the most widely used set of summary figures is known as measures of location."

**Characteristics for a good or an ideal average :**

The following properties should possess for an ideal average.

1. It should be rigidly defined.
2. It should be easy to understand and compute.
3. It should be based on all items in the data.
4. Its definition shall be in the form of a mathematical formula.
5. It should be capable of further algebraic treatment.
6. It should have sampling stability.
7. It should be capable of being used in further statistical computations or processing.

Besides the above requisites, a good average should represent maximum characteristics of the data, its value should be nearest to the most items of the given series.

**Arithmetic mean or mean :**

Arithmetic mean or simply the mean of a variable is defined as the sum of the observations divided by the number of observations. If the variable x assumes n values $x_1$, $x_2$ ..$x_n$ then the mean, $\bar{x}$, is given by

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n}$$

$$= \frac{1}{n}\sum_{i=1}^{n} x_i$$

This formula is for the ungrouped or raw data.

**Example 1 :**

Calculate the mean for 2, 4, 6, 8, 10

**Solution:**

$$\bar{x} = \frac{2+4+6+8+10}{5}$$

$$= \frac{30}{5} = 6$$

**Short-Cut method :** (Deviation Method)

Under this method an assumed or an arbitrary average (indicated by A) is used as the basis of calculation of deviations from individual values. The formula is

$$\bar{x} = A + \frac{\sum d}{n}$$

where, A = the assumed mean or any value in x

d = the deviation of each value from the assumed mean

**Example 2 :**

A student's marks in 5 subjects are 75, 68, 80, 92, 56. Find his average mark.

**Solution:**

| X | d=x-A |
|---|---|
| 75 | 7 |
| A $\boxed{68}$ | 0 |
| 80 | 12 |
| 92 | 24 |
| 56 | -12 |
| Total | 31 |

0
-7
-5
-17
17

-12
-17
-29

$$\bar{x} = A + \frac{\sum d}{n}$$

$$= 68 + \frac{31}{5}$$

$$= 68 + 6.2$$

$$= 74.2$$

$75 - \frac{10}{5}$
$= 75 - 2$
$= 73.0$

**Grouped Data :**

The mean for grouped data is obtained from the following formula:

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

where $x$ = the mid-point of individual class
$f$ = the frequency of individual class
N = the sum of the frequencies or total frequencies.

**Short-cut method :**

$$\bar{x} = A + \frac{\sum fd}{N} \times c$$

where $d = \frac{x - A}{c}$

A = any value in x
N = total frequency
c = width of the class interval

**Example 3:**

Given the following frequency distribution, calculate the arithmetic mean

| | | | | | | |
|---|---|---|---|---|---|---|
| Marks | : 64 | 63 | 62 | 61 | 60 | 59 |
| Number of Students | : 8 | 18 | 12 | 9 | 7 | 6 |

**Solution:**

| X | F | fx | d=x-A | fd |
|---|---|---|---|---|
| 64 | 8 | 512 | 2 | 16 |
| 63 | 18 | 1134 | 1 | 18 |
| A $\boxed{62}$ | 12 | 744 | 0 | 0 |
| 61 | 9 | 549 | –1 | –9 |
| 60 | 7 | 420 | –2 | –14 |
| 59 | 6 | 354 | –3 | –18 |
| | 60 | 3713 | | - 7 |

**Direct method**

$$\bar{x} = \frac{\sum fx}{N} = \frac{3713}{60} = 61.88$$

**Short-cut method**

$$\bar{x} = A + \frac{\sum fd}{N} = 62 - \frac{7}{60} = 61.88$$

**Example 4 :**

Following is the distribution of persons according to different income groups. Calculate arithmetic mean.

| Income Rs(100) | 0-10 | 10-20 | 20-30 | 30-40 | 40-50 | 50-60 | 60-70 |
|---|---|---|---|---|---|---|---|
| Number of persons | 6 | 8 | 10 | 12 | 7 | 4 | 3 |

**Solution:**

| Income C.I | Number of Persons (f) | Mid X | $d = \frac{x-A}{c}$ | Fd |
|---|---|---|---|---|
| 0-10 | 6 | 5 | -3 | -18 |
| 10-20 | 8 | 15 | -2 | -16 |
| 20-30 | 10 | 25 | -1 | -10 |
| 30-40 | 12 | A $\boxed{35}$ | 0 | 0 |
| 40-50 | 7 | 45 | 1 | 7 |
| 50-60 | 4 | 55 | 2 | 8 |
| 60-70 | 3 | 65 | 3 | 9 |
| | N = 50 | | | -20 |

$$\text{Mean} = \bar{x} = A + \frac{\sum fd}{N} \times C$$

$$= 35 - \frac{20}{50} \times 10$$

$$= 35 - 4$$

$$= 31$$

**Merits and demerits of Arithmetic mean :**

**Merits:**

1. It is rigidly defined. (same result obtained by using any method
2. It is easy to understand and easy to calculate.
3. If the number of items is sufficiently large, it is more accurate and more reliable.
4. It is a calculated value and is not based on its position in the series.
5. It is possible to calculate even if some of the details of the data are lacking.
6. Of all averages, it is affected least by fluctuations of sampling.
7. It provides a good basis for comparison.

**Demerits:**

1. It cannot be obtained by inspection nor located through a frequency graph.
2. It cannot be in the study of qualitative phenomena not capable of numerical measurement i.e., Intelligence, beauty, honesty etc.,
3. It can ignore any single item only at the risk of losing its accuracy.
4. It is affected very much by extreme values.
5. It cannot be calculated for open-end classes.
6. It may lead to fallacious conclusions, if the details of the data from which it is computed are not given.

**Weighted Arithmetic mean :**

For calculating simple mean, we suppose that all the values or the sizes of items in the distribution have equal importance. But, in practical life this may not be so. In case some items are more

Solution:

| Designation | Monthly salary,x | Strength of the cadre,w | wx |
|---|---|---|---|
| Class 1 officer | 1,500 | 10 | 15,000 |
| Class 2 officer | 800 | 20 | 16,000 |
| Subordinate staff | 500 | 70 | 35,000 |
| Clerical staff | 250 | 100 | 25,000 |
| Lower staff | 100 | 150 | 15,000 |
| | | 350 | 1,06,000 |

Weighted average, $\bar{x}_w = \dfrac{\sum wx}{\sum w}$

$$= \frac{106000}{350}$$

$$= \text{Rs. } 302.86$$

**Harmonic mean (H.M) :**

Harmonic mean of a set of observations is defined as the reciprocal of the arithmetic average of the reciprocal of the given values. If $x_1, x_2, \ldots x_n$ are n observations,

$$H.M = \frac{n}{\sum_{i=1}^{n}\left(\frac{1}{x_i}\right)}$$

For a frequency distribution

$$H.M = \frac{N}{\sum_{i=1}^{n} f\left(\frac{1}{x_i}\right)}$$

**Example 6:**

From the given data calculate H.M 5,10,17,24,30

| X | $\frac{1}{x}$ |
|---|---|
| 5 | 0.2000 |
| 10 | 0.1000 |
| 17 | 0.0588 |
| 24 | 0.0417 |
| 30 | 0.0333 |
| Total | 0.4338 |

$$\text{H.M} = \frac{n}{\sum\left[\frac{1}{x}\right]}$$

$$= \frac{5}{0.4338} = 11.526$$

$= \frac{50000}{4338}$

$= 11.526$

**Example 7:**

The marks secured by some students of a class are given below. Calculate the harmonic mean.

| Marks | 20 | 21 | 22 | 23 | 24 | 25 |
|---|---|---|---|---|---|---|
| Number of Students | 4 | 2 | 7 | 1 | 3 | 1 |

**Solution:**

| Marks $X$ | No of students f | $\frac{1}{x}$ | $f(\frac{1}{x})$ |
|---|---|---|---|
| 20 | 4 | 0.0500 | 0.2000 |
| 21 | 2 | 0.0476 | 0.0952 |
| 22 | 7 | 0.0454 | 0.3178 |
| 23 | 1 | 0.0435 | 0.0435 |
| 24 | 3 | 0.0417 | 0.1251 |
| 25 | 1 | 0.0400 | 0.0400 |
|  | 18 |  | 0.8216 |

$$H.M = \frac{N}{\sum f\left[\frac{1}{x}\right]}$$

$$= \frac{\cancel{18}}{\cancel{0.1968}} = \cancel{21.91} \qquad \frac{18}{0.8216} = 21.91$$

**Merits of H.M :**

1. It is rigidly defined.
2. It is defined on all observations.
3. It is amenable to further algebraic treatment.
4. It is the most suitable average when it is desired to give greater weight to smaller observations and less weight to the larger ones.

**Demerits of H.M :**

1. It is not easily understood.
2. It is difficult to compute.
3. It is only a summary figure and may not be the actual item in the series
4. It gives greater importance to small items and is therefore, useful only when small items have to be given greater weightage.

**Geometric mean :**

The geometric mean of a series containing n observations is the $n^{th}$ root of the product of the values. If $x_1, x_2 .., x_n$ are observations then

$$G.M = \sqrt[n]{x_1 . x_2 ... x_n}$$

$$= (x_1.x_2 .. x_n)^{1/n}$$

$$\log GM = \frac{1}{n} \log(x_1.x_2 .. x_n)$$

$$= \frac{1}{n} (\log x_1 + \log x_2 + .. + \log x_n$$

$$= \frac{\sum \log x_i}{n}$$

$$GM = \text{Antilog } \frac{\sum \log x_i}{n}$$

$$= \frac{1}{0.0009}$$



$$= \frac{10000}{9}$$

$$\frac{100000}{12467329}$$

$$= 0.008$$

For grouped data

$$GM = \text{Antilog}\left[\frac{\sum f \log x_i}{N}\right]$$

**Example 8:**

Calculate the geometric mean of the following series of monthly income of a batch of families 180,250,490,1400,1050

| x | logx |
|---|---|
| 180 | 2.2553 |
| 250 | 2.3979 |
| 490 | 2.6902 |
| 1400 | 3.1461 |
| 1050 | 3.0212 |
| | 13.5107 |

$$GM = \text{Antilog}\left[\frac{\sum \log x}{n}\right]$$

$$= \text{Antilog}\ \frac{13.5107}{5}$$

$$= \text{Antilog}\ 2.7021 \quad = 503.6$$

**Example 9:**

Calculate the average income per head from the data given below. Use geometric mean.

| Class of people | Number of families | Monthly income per head (Rs) |
|---|---|---|
| Landlords | 2 | 5000 |
| Cultivators | 100 | 400 |
| Landless – labours | 50 | 200 |
| Money – lenders | 4 | 3750 |
| Office Assistants | 6 | 3000 |
| Shop keepers | 8 | 750 |
| Carpenters | 6 | 600 |
| Weavers | 10 | 300 |

Solution:

| Class of people | Annual income ( Rs) X | Number of families (f) | Log x | f logx |
|---|---|---|---|---|
| Landlords | 5000 | 2 | 3.6990 | 7.398 |
| Cultivators | 400 | 100 | 2.6021 | 260.210 |
| Land less – labours | 200 | 50 | 2.3010 | 115.050 |
| Money – lenders | 3750 | 4 | 3.5740 | 14.296 |
| Office Assistants | 3000 | 6 | 3.4771 | 20.863 |
| Shop keepers | 750 | 8 | 2.8751 | 23.2008 |
| Carpenters | 600 | 6 | 2.7782 | 16.669 |
| Weavers | 300 | 10 | 2.4771 | 24.771 |
| | | 186 | | 482.257 |

$$GM = \text{Antilog}\left[\frac{\sum f \log x}{N}\right]$$

$$= \text{Antilog}\left[\frac{482.257}{186}\right]$$

$$= \text{Antilog}\ (2.5928)$$

$$= \text{Rs } 391.50$$

**Merits of Geometric mean :**

1. It is rigidly defined
2. It is based on all items
3. It is very suitable for averaging ratios, rates and percentages
4. It is capable of further mathematical treatment.
5. Unlike AM, it is not affected much by the presence of extreme values

**Demerits of Geometric mean:**

1. It cannot be used when the values are negative or if any of the observations is zero
2. It is difficult to calculate particularly when the items are very large or when there is a frequency distribution.

$$P_{53} = l + \frac{\frac{53N}{100} - m}{f} \times c$$

$$= 20 + \frac{41.34 - 41}{20} \times 5$$

$$= 20 + 0.085 = 20.085.$$

**Mode :**

The mode refers to that value in a distribution, which occur most frequently. It is an actual value, which has the highest concentration of items in and around it.

According to Croxton and Cowden " The mode of a distribution is the value at the point around which the items tend to be most heavily concentrated. It may be regarded at the most typical of a series of values".

It shows the centre of concentration of the frequency in around a given value. Therefore, where the purpose is to know the point of the highest concentration it is preferred. It is, thus, a positional measure.

Its importance is very great in marketing studies where a manager is interested in knowing about the size, which has the highest concentration of items. For example, in placing an order for shoes or ready-made garments the modal size helps because this sizes and other sizes around in common demand.

**Computation of the mode:**

**Ungrouped or Raw Data:**

For ungrouped data or a series of individual observations, mode is often found by mere inspection.

**Example 29:**

2 , 7, 10, 15, 10, 17, 8, 10, 2

$\therefore$ Mode = $M_0 = 10$

In some cases the mode may be absent while in some cases there may be more than one mode.

* Mode (grouped data)

$$\text{Mode} = L_1 + \frac{(L_2 - L_1)(f_1 - f_0)}{2f_1 - f_0 - f_2}$$

* Steps : Find the class interval which has maximum frequency.
2) Then use interpulation formula to find the exact model value.
3) In this formula $l_1$ = lower limit of the model class.

$l_2$ = Upper limit of the model class.

$f_1$ = Frequency of the model class.

$f_0$ = frequency of the class preceding model class.

$f_2$ = Frequency of the class succeeding model class.

eg:

* Calculate the mode from the following data.

| Daily Wages | 10-20 | 20-30 | 30-40 | 40-50 | 50-60 | 60-70 |
|---|---|---|---|---|---|---|
| No. of workers | 5 | 6 | 10 | $\boxed{16}$ | 12 | 8 |

$\therefore$ N =

| Daily wages | No of workers |
|---|---|
| 10-20 | 5 |
| 20-30 | 6 |
| 30-40 | 10 $f_0$ |
| 40-50 | 16 $f_1$ |
| 50-60 | 12 $f_2$ |
| 60-70 | 8 |

~~L = 40~~

It may be noted that 16 is the highest frequency and model class is 40 - 50

$L_1 = 40$

$L_2 = 50$

$f_1 = 16$

$f_0 = 10$

$f_2 = 12$.

$$\text{Mode} = \frac{L_1 + (L_2 - L_1)(f_1 - f_0)}{2f_1 - f_0 - f_2}$$

$$= 40 + \left( \frac{(50-40)\ (16-10)}{2 \times 16 - 10 - 12} \right)$$

$$= 40 + \left( \frac{10 \times 6}{32 - 22} \right)$$

$$= 40 + \frac{60}{10}$$

$$= 46$$

Mode = 46 .

that the median and mode are called the positional measures of an average.

**Median :**

The median is that value of the variate which divides the group into two equal parts, one part comprising all values greater, and the other, all values less than median.

**Ungrouped or Raw data :**

Arrange the given values in the increasing or decreasing order. If the number of values are odd, median is the middle value .If the number of values are even, median is the mean of middle two values.

By formula

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

**Example 11:**

When odd number of values are given. Find median for the following data

25, 18, 27, 10, 8, 30, 42, 20, 53

**Solution:**

Arranging the data in the increasing order 8, 10, 18, 20, 25, 27, 30, 42, 53

The middle value is the $5^{\text{th}}$ item i.e., 25 is the median

Using formula

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

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

$$= \left(\frac{10}{2}\right)^{\text{th}} \text{item}$$

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

$$= 25$$

**Example 12 :**

When even number of values are given. Find median for the following data

5, 8, 12, 30, 18, 10, 2, 22

**Solution:**

Arranging the data in the increasing order 2, 5, 8, 10, 12, 18, 22, 30

Here median is the mean of the middle two items (ie) mean of (10,12) ie

$$= \left(\frac{10+12}{2}\right) = 11$$

$$\therefore \text{median} = 11.$$

Using the formula

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

2

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

$$= \left(\frac{9}{2}\right)^{\text{th}} \text{item} = 4.5^{\text{th}} \text{item}$$

$$= 4^{\text{th}} \text{item} + \left(\frac{1}{2}\right)(5^{\text{th}} \text{item} - 4^{\text{th}} \text{item})$$

$4 + \frac{1}{2}(1)$

$= 4 + 0.5$

$= 4.5$

$$= 10 + \left(\frac{1}{2}\right)[12-10]$$

$$= 10 + \left(\frac{1}{2}\right) \times 2$$

$$= 10 + 1$$

$$= 11$$

**Example 13:**

The following table represents the marks obtained by a batch of 10 students in certain class tests in statistics and Accountancy.

| Serial No | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Marks (Statistics) | 53 | 55 | 52 | 32 | 30 | 60 | 47 | 46 | 35 | 28 |
| Marks (Accountancy) | 57 | 45 | 24 | 31 | 25 | 84 | 43 | 80 | 32 | 72 |

Indicate in which subject is the level of knowledge higher ?

**Solution:**

For such question, median is the most suitable measure of central tendency. The mark in the two subjects are first arranged in increasing order as follows:

| Serial No | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Marks in Statistics | 28 | 30 | 32 | 35 | 46 | 47 | 52 | 53 | 55 | 60 |
| Marks in Accountancy | 24 | 25 | 31 | 32 | 43 | 45 | 57 | 72 | 80 | 84 |

$$\text{Median} = \left(\frac{n+1}{2}\right)^{\text{th}} \text{item} = \left(\frac{10+1}{2}\right)^{\text{th}} \text{item} = 5.5^{\text{th}} \text{ item}$$

$$= \frac{Value\ of\ 5^{th}\ item + value\ of\ 6^{th}\ item}{2}$$

$$\text{Md (Statistics)} = \frac{46+47}{2} = 46.5$$

$$\text{Md (Accountancy)} = \frac{43+45}{2} = 44$$

There fore the level of knowledge in Statistics is higher than that in Accountancy.

**Grouped Data:**

In a grouped distribution, valucs arc associated with frequencies. Grouping can be in the form of a discrete frequency distribution or a continuous frequency distribution. Whatever may be the type of distribution , cumulative frequencies have to be calculated to know the total number of items.

**Cumulative frequency : (cf)**

Cumulative frequency of each class is the sum of the frequency of the class and the frequencies of the pervious classes, ie adding the frequencies successively, so that the last cumulative frequency gives the total number of items.

**Discrete Series:**

Step1: Find cumulative frequencies.

Step2: Find $\left(\dfrac{N+1}{2}\right)$

Step3: See in the cumulative frequencies the value just greater than $\left(\dfrac{N+1}{2}\right)$

Step4: Then the corresponding value of x is median.

**Example 14:**

The following data pertaining to the number of members in a family. Find median size of the family.

| Number of members **x** | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Frequency F | 1 | 3 | 5 | 6 | 10 | 13 | 9 | 5 | 3 | 2 | 2 | 1 |

**Solution:**

| X | f | cf |
|---|---|---|
| 1 | 1 | 1 |
| 2 | 3 | 4 |
| 3 | 5 | 9 |
| 4 | 6 | 15 |
| 5 | 10 | 25 |
| 6 | 13 | 38 |
| 7 | 9 | 47 |
| 8 | 5 | 52 |
| 9 | 3 | 55 |
| 10 | 2 | 57 |
| 11 | 2 | 59 |
| 12 | 1 | 60 |
| | 60 | |

Median = size of $\left(\dfrac{N+1}{2}\right)^{th}$ item

$$= \text{size of} \left(\frac{60+1}{2}\right)^{th} \text{item}$$

$$= 30.5^{th} \text{ item}$$

The cumulative frequencies just greater than 30.5 is 38.and the value of x corresponding to 38 is 6.Hence the median size is 6 members per family.

**Note:**

It is an appropriate method because a fractional value given by mean does not indicate the average number of members in a family.

**Continuous Series:**

The steps given below are followed for the calculation of median in continuous series.

Step1: Find cumulative frequencies.

Step2: Find $\left(\frac{N}{2}\right)$

Step3: See in the cumulative frequency the value first greater than $\left(\frac{N}{2}\right)$, Then the corresponding class interval is called the Median class. Then apply the formula

$$\text{Median} = l + \frac{\frac{N}{2} - m}{f} \times c$$

Where $l$ = Lower limit of the median class

m = cumulative frequency preceding the median

c = width of the median class

f = frequency in the median class.

N = Total frequency.

**Note :**

If the class intervals are given in inclusive type convert them into exclusive type and call it as true class interval and consider lower limit in this.

**Example 15:**

The following table gives the frequency distribution of 325 workers of a factory, according to their average monthly income in a certain year.

| Income group (in Rs) | Number of workers |
|---|---|
| Below 100 | 1 |
| 100-150 | 20 |
| 150-200 | 42 |
| 200-250 | 55 |
| 250-300 | 62 |
| 300-350 | 45 |
| 350-400 | 30 |
| 400-450 | 25 |
| 450-500 | 15 |
| 500-550 | 18 |
| 550-600 | 10 |
| 600 and above | 2 |
| | 325 |

Calculate median income

**Solution:**

| Income group (Class-interval) | Number of workers (Frequency) | Cumulative frequency c.f |
|---|---|---|
| Below 100 | 1 | 1 |
| 100-150 | 20 | 21 |
| 150-200 | 42 | 63 |
| 200-250 | 55 | 118 |
| 250-300 | 62 | 180 |
| 300-350 | 45 | 225 |
| 350-400 | 30 | 255 |
| 400-450 | 25 | 280 |
| 450-500 | 15 | 295 |
| 500-550 | 18 | 313 |
| 550-600 | 10 | 323 |
| 600 and above | 2 | 325 |
| | 325 | |

$$\frac{N}{2} = \frac{325}{2} = 162.5$$

Here $l = 250$, N = 325, f = 62, c = 50, m = 118

$$Md = 250 + \left(\frac{162.5 - 118}{62}\right) \times 50$$

$$= 250 + 35.89$$

$$= 285.89$$

**Example 16:**

Calculate median from the following data

| Value | 0-4 | 5-9 | 10-14 | 15-19 | 20-24 | 25-29 | 30-34 | 35-39 |
|---|---|---|---|---|---|---|---|---|
| Frequency | 5 | 8 | 10 | 12 | 7 | 6 | 3 | 2 |

| Value | f | True class interval | c.f. |
|---|---|---|---|
| 0-4 | 5 | 0.5-4.5 | 5 |
| 5-9 | 8 | 4.5-9.5 | 13 |
| 10-14 | 10 | 9.5-14.5 | 23 |
| 15-19 | 12 | 14.5-19.5 | 35 |
| 20-24 | 7 | 19.5-24.5 | 42 |
| 25-29 | 6 | 24.5-29.5 | 48 |
| 30-34 | 3 | 29.5-34.5 | 51 |
| 35-39 | 2 | 34.5-39.5 | 53 |
| | 53 | | |

$$\left(\frac{N}{2}\right) = \left(\frac{53}{2}\right) = 26.5$$

$$Md = l + \frac{\frac{N}{2} - m}{f} \times c$$

$$= 14.5 + \frac{26.5 - 23}{12} \times \cancel{5}\,4$$

$$= 14.5 + 1.46 = 15.96$$ 15.66

**Example 17:**

Following are the daily wages of workers in a textile. Find the median.

| Wages ( in Rs.) | Number of workers |
|---|---|
| less than 100 | 5 |
| less than 200 | 12 |
| less than 300 | 20 |
| less than 400 | 32 |
| less than 500 | 40 |
| less than 600 | 45 |
| less than 700 | 52 |
| less than 800 | 60 |
| less than 900 | 68 |
| less than 1000 | 75 |

**Solution :**

We are given upper limit and less than cumulative frequencies. First find the class-intervals and the frequencies. Since the values are increasing by 100, hence the width of the class interval equal to 100.

| Class interval | f | c.f |
|---|---|---|
| 0-100 | 5 | 5 |
| 100-200 | 7 | 12 |
| 200-300 | 8 | 20 |
| 300- 400 | 12 | 32 |
| 400-500 | 8 | 40 |
| 500-600 | 5 | 45 |
| 600-700 | 7 | 52 |
| 700-800 | 8 | 60 |
| 800-900 | 8 | 68 |
| 900-1000 | 7 | 75 |
| | 75 | |

$$\left(\frac{N}{2}\right) = \left(\frac{75}{2}\right) = 37.5$$

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

$$= 400 + \left(\frac{37.5 - 32}{8}\right) \times 100 = 400 + 68.75 = 468.75$$

**Example 18:**

Find median for the data given below.

| Marks | Number of students |
|---|---|
| Greater than 10 | 70 |
| Greater than 20 | 62 |
| Greater than 30 | 50 |
| Greater than 40 | 38 |
| Greater than 50 | 30 |
| Greater than 60 | 24 |
| Greater than 70 | 17 |
| Greater than 80 | 9 |
| Greater than 90 | 4 |

**Solution :**

Here we are given lower limit and more than cumulative frequencies.

| Class interval | f | More than c.f | Less than c.f |
|---|---|---|---|
| 10-20 | 8 | 70 | 8 |
| 20-30 | 12 | 62 | 20 |
| 30-40 | 12 | 50 | 32 |
| 40-50 | 8 | 38 | 40 |
| 50-60 | 6 | 30 | 46 |
| 60-70 | 7 | 24 | 53 |
| 70-80 | 8 | 17 | 61 |
| 80-90 | 5 | 9 | 66 |
| 90-100 | 4 | 4 | 70 |
| | 70 | | |

$$\left(\frac{N}{2}\right) = \left(\frac{70}{2}\right) = 35$$

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

$$= 40 + \left( \frac{35 - 32}{8} \right) \times 10$$

$$= 40 + 3.75$$

$$= 43.75$$

**Example 19:**

Compute median for the following data.

| Mid-Value | 5 | 15 | 25 | 35 | 45 | 55 | 65 | 75 |
|---|---|---|---|---|---|---|---|---|
| Frequency | 7 | 10 | 15 | 17 | 8 | 4 | 6 | 7 |

**Solution :**

Here values in multiples of 10, so width of the class interval is 10.

| Mid x | C.I | f | c.f |
|---|---|---|---|
| 5 | 0-10 | 7 | 7 |
| 15 | 10-20 | 10 | 17 |
| 25 | 20-30 | 15 | 32 |
| 35 | 30-40 | 17 | 49 |
| 45 | 40-50 | 8 | 57 |
| 55 | 50-60 | 4 | 61 |
| 65 | 60-70 | 6 | 67 |
| 75 | 70-80 | 7 | 74 |
| | | 74 | |

$$\left( \frac{N}{2} \right) = \left( \frac{74}{2} \right) = 37$$

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

$$= 30 + \left(\frac{37-32}{17}\right) \times 10$$
$$= 30 + 2.94$$
$$= 32.94$$

**Graphic method for Location of median:**

Median can be located with the help of the cumulative frequency curve or ' ogive' . The procedure for locating median in a grouped data is as follows:

Step1: The class boundaries, where there are no gaps between consecutive classes, are represented on the horizontal axis (x-axis).

Step2: The cumulative frequency corresponding to different classes is plotted on the vertical axis (y-axis) against the upper limit of the class interval (or against the variate value in the case of a discrete series.)

Step3: The curve obtained on joining the points by means of freehand drawing is called the ' ogive'. The ogive so drawn may be either a (i) less than ogive or a (ii) more than ogive.

Step4: The value of $\frac{N}{2}$ or $\frac{N+1}{2}$ is marked on the y-axis, where N is the total frequency.

Step5: A horizontal straight line is drawn from the point $\frac{N}{2}$ or $\frac{N+1}{2}$ on the y-axis parallel to x-axis to meet the ogive.

Step6: A vertical straight line is drawn from the point of intersection perpendicular to the horizontal axis.

Step7: The point of intersection of the perpendicular to the x-axis gives the value of the median.

**Remarks :**

1. From the point of intersection of ' less than' and ' more than' ogives, if a perpendicular is drawn on the x-axis, the point so obtained on the horizontal axis gives the value of the median.
2. If ogive is drawn using cumulated percentage frequencies, then we draw a straight line from the point intersecting 50

**Merits of Median :**

1. Median is not influenced by extreme values because it is a positional average.
2. Median can be calculated in case of distribution with open-end intervals.
3. Median can be located even if the data are incomplete.
4. Median can be located even for qualitative factors such as ability, honesty etc.

**Demerits of Median :**

1. A slight change in the series may bring drastic change in median value.
2. In case of even number of items or continuous series, median is an estimated value other than any value in the series.
3. It is not suitable for further mathematical treatment except its use in mean deviation.
4. It is not taken into account all the observations.

**Quartiles :**

The quartiles divide the distribution in four parts. There are three quartiles. The second quartile divides the distribution into two halves and therefore is the same as the median. The first (lower) quartile ($Q_1$) marks off the first one-fourth, the third (upper) quartile ($Q_3$) marks off the three-fourth.

**Raw or ungrouped data:**

First arrange the given data in the increasing order and use the formula for $Q_1$ and $Q_3$ then quartile deviation, Q.D is given by

$$Q.D = \frac{Q_3 - Q_1}{2}$$

Where $Q_1 = \left(\frac{n+1}{4}\right)^{th}$ item and $Q_3 = 3\left(\frac{n+1}{4}\right)^{th}$ item

**Example 22 :**

Compute quartiles for the data given below 25,18,30, 8, 15, 5, 10, 35, 40, 45

**Solution :**

5, 8, 10, 15, 18,25, 30,35,40, 45

$$Q_1 = \left(\frac{n+1}{4}\right)^{th} \text{ item}$$

$$= \left(\frac{10+1}{4}\right)^{th} \text{ item}$$

$$= (2.75)^{th} \text{ item}$$

$$= 2^{nd} \text{ item} + \left(\frac{3}{4}\right) (3^{rd} \text{ item} - 2^{nd} \text{ item})$$

$$= 8 + \frac{3}{4} \quad (10-8)$$

$$= 8 + \frac{3}{4} \times 2$$

$$= 8 + 1.5$$

$$= 9.5$$

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

$$= 3 \times (2.75)^{th} \text{ item}$$

$$= (8.25)^{th} \text{ item}$$

$$= 8^{th} \text{ item} + \frac{1}{4} \; [9^{th} \text{ item} - 8^{th} \text{ item}]$$

$$= 35 + \frac{1}{4} [40-35]$$

$$= 35 + 1.25 = 36.25$$

**Discrete Series :**

Step1: Find cumulative frequencies.

Step2: Find $\left(\frac{N+1}{4}\right)$

Step3: See in the cumulative frequencies , the value just greater than $\left(\frac{N+1}{4}\right)$ ,then the corresponding value of $x$ is $Q_1$

Step4: Find $3\left(\frac{N+1}{4}\right)$

Step5: See in the cumulative frequencies, the value just greater than $3\left(\frac{N+1}{4}\right)$ ,then the corresponding value of x is $Q_3$

**Example 23:**

Compute quartiles for the data given bellow.

| X | 5 | 8 | 12 | 15 | 19 | 24 | 30 |
|---|---|---|---|---|---|---|---|
| f | 4 | 3 | 2 | 4 | 5 | 2 | 4 |

**Solution:**

| x | f | c.f |
|---|---|---|
| 5 | 4 | 4 |
| 8 | 3 | 7 |
| 12 | 2 | 9 |
| 15 | 4 | 13 |
| 19 | 5 | 18 |
| 24 | 2 | 20 |
| 30 | 4 | 24 |
| Total | 24 | |

$Q_1 = \left(\frac{N+1}{4}\right)^{th}$ item $= \left(\frac{24+1}{4}\right) = \left(\frac{25}{4}\right) = 6.25^{th}$ item

$Q_3 = 3\left(\frac{N+1}{4}\right)^{th}$ item $= 3\left(\frac{24+1}{4}\right) = 18.75^{th}$ item $\therefore Q_1 = 8;\ Q_3 = 24$

**Continuous series :**

Step1: Find cumulative frequencies

Step2: Find $\left(\frac{N}{4}\right)$

Step3: See in the cumulative frequencies, the value just greater than $\left(\frac{N}{4}\right)$ , then the corresponding class interval is called first quartile class.

Step4: Find $3\left(\frac{N}{4}\right)$ See in the cumulative frequencies the value just greater than $3\left(\frac{N}{4}\right)$ then the corresponding class interval is called 3rd quartile class. Then apply the respective formulae

$$Q_1 = l_1 + \frac{\frac{N}{4} - m_1}{f_1} \times c_1$$

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

Where $l_1$ = lower limit of the first quartile class
$f_1$ = frequency of the first quartile class
$c_1$ = width of the first quartile class
$m_1$ = c.f. preceding the first quartile class
$l_3$ = lower limit of the 3rd quartile class
$f_3$ = frequency of the 3rd quartile class
$c_3$ = width of the 3rd quartile class
$m_3$ = c.f. preceding the 3rd quartile class

**Example 24:**

The following series relates to the marks secured by students in an examination.

| Marks | No. of students |
|---|---|
| 0-10 | 11 |
| 10-20 | 18 |
| 20-30 | 25 |
| 30-40 | 28 |
| 40-50 | 30 |
| 50-60 | 33 |
| 60-70 | 22 |
| 70-80 | 15 |
| 80-90 | 12 |
| 90-100 | 10 |

Find the quartiles

**Solution :**

| C.I. | f | cf |
|---|---|---|
| 0-10 | 11 | 11 |
| 10-20 | 18 | 29 $m_1$ |
| 20-30 | 25 $f_1$ | 54 |
| 30-40 | 28 | 82 |
| 40-50 | 30 | 112 |
| 50-60 | 33 | 145 $m_3$ |
| 60-70 | $f_3$ 22 | 167 |
| 70-80 | 15 | 182 |
| 80-90 | $l_3$ 12 | 194 |
| 90-100 | 10 | 204 |
| | 204 | |

$l_1$ $c_1 = 10$

$$\left(\frac{N}{4}\right) = \left(\frac{204}{4}\right) = 51 \qquad 3\left(\frac{N}{4}\right) = 153$$

$$Q_1 = l_1 + \frac{\frac{N}{4} - m_1}{f_1} \times c_1$$

$$= 20 + \frac{51-29}{25} \times 10 = 20 + 8.8 = 28.8$$

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

$$= 60 + \frac{153-145}{22} \times 12 = 60 + 4.36 = \cancel{64.36} \quad 63.88$$

**Deciles :**

These are the values, which divide the total number of observation into 10 equal parts. These are 9 deciles $D_1, D_2 \ldots D_9$. These are all called first decile, second decile..etc.,

## Deciles for Raw data or ungrouped data

**Example 25:**

Compute $D_5$ for the data given below

5, 24, 36, 12, 20, 8

**Solution :**

Arranging the given values in the increasing order

5, 8, 12, 20, 24, 36

$$D_5 = \left(\frac{5(n+1)}{10}\right)^{th} \text{ observation}$$

$$= \left(\frac{5(6+1)}{10}\right)^{th} \text{ observation}$$

$$= (3.5)^{th} \text{ observation}$$

$$= 3^{rd} \text{ item} + \frac{1}{2}[4^{th} \text{ item} - 3^{rd} \text{ item}]$$

$$= 12 + \frac{1}{2}[20 - 12] \quad = 12 + 4 \quad = 16$$

**Deciles for Grouped data :**

**Example 26:**

Calculate $D_3$ and $D_7$ for the data given below

| Class Interval | 0-10 | 10-20 | 20-30 | 30-40 | 40-50 | 50-60 | 60-70 |
|---|---|---|---|---|---|---|---|
| Frequency : | 5 | 7 | 12 | 16 | 10 | 8 | 4 |

**Solution :**

| C.I | f | c.f |
|---|---|---|
| 0-10 | 5 | 5 |
| 10-20 | 7 | 12 |
| 20-30 | 12 | 24 |
| 30-40 | 16 | 40 |
| 40-50 | 10 | 50 |
| 50-60 | 8 | 58 |
| 60-70 | 4 | 62 |
| | 62 | |

$$D_3 \text{ item} = \left(\frac{3N}{10}\right)^{th} \text{item}$$

$$= \left(\frac{3\times 62}{10}\right)^{th} \text{item}$$

$$= (18.6)^{th} \text{ item}$$

which lies in the interval 20-30

$$\therefore D_3 = l + \frac{3\left(\frac{N}{10}\right) - m}{f} \times c$$

$$= 20 + \frac{18.6 - 12}{12} \times 10$$

$$= 20 + 5\cancel{7}5 = 25.5 \quad \boxed{25.75}$$

$$D_7 \text{ item} = \left(\frac{7\times N}{10}\right)^{th} \text{item}$$

$$= \left(\frac{7\times 62}{10}\right)^{th} \text{item}$$

$$\frac{441}{10} = \left(\frac{\cancel{434}}{10}\right)^{th} \text{item} = \cancel{(43.4)^{th}} \text{ item} \quad 44.1$$

which lies in the interval(40-50)

$$D_7 = l + \frac{\left(\frac{7N}{10}\right) - m}{f} \times c$$

$$= 40 + \frac{43.4 - 40}{10} \times 10$$

$$= 40 + 3.4 = \cancel{43.4} \quad 44.1$$

**Percentiles :**

The percentile values divide the distribution into 100 parts each containing 1 percent of the cases. The percentile ($P_k$) is that value of the variable up to which lie exactly k% of the total number of observations.

**Relationship :**

$$P_{25} = Q_1 \; ; \; P_{50} = D_5 = Q_2 = \text{Median and } P_{75} = Q_3$$

**Percentile for Raw Data or Ungrouped Data :**

**Example 27:**

Calculate $P_{15}$ for the data given below:

5, 24 , 36 , 12 , 20 , 8

Arranging the given values in the increasing order.

5, 8, 12, 20, 24, 36

$$P_{15} = \left(\frac{15(n+1)}{100}\right)^{th} \text{ item}$$

$$= \left(\frac{15 \times 7}{100}\right)^{th} \text{ item}$$

$$= (1.05)^{th} \text{ item}$$

$$= 1^{st} \text{ item} + 0.05\,(2^{nd} \text{ item} - 1^{st} \text{ item})$$

$$= 5 + 0.05\,(8-5)$$

$$= 5 + 0.15 = 5.15$$

**Percentile for grouped data :**

**Example 28:**

Find $P_{53}$ for the following frequency distribution.

| Class interval | 0-5 | 5-10 | 10-15 | 15-20 | 20-25 | 25-30 | 30-35 | 35-40 |
|---|---|---|---|---|---|---|---|---|
| Frequency | 5 | 8 | 12 | 16 | 20 | 10 | 4 | 3 |

**Solution:**

| Class Interval | Frequency | C.f |
|---|---|---|
| 0-5 | 5 | 5 |
| 5-10 | 8 | 13 |
| 10-15 | 12 | 25 |
| 15-20 | 16 | 41 |
| 20-25 | 20 | 61 |
| 25-30 | 10 | 71 |
| 30-35 | 4 | 75 |
| 35-40 | 3 | 78 |
| Total | 78 | |

$$P_{53} = l + \frac{\frac{53N}{100} - m}{f} \times c$$

$$= 20 + \frac{41.34 - 41}{20} \times 5$$

$$= 20 + 0.085 \quad = 20.085.$$

## Measures of Central Tendency:

In the study of a population with respect to one in which we are interested we may get a large number of observations. It is not possible to grasp any idea about the characteristic when we look at all the observations. So it is better to get one number for one group. That number must be a good representative one for all the observations to give a clear picture of that characteristic. Such representative number can be a central value for all these observations. This central value is called a measure of central tendency or an average or a measure of locations. There are five averages. Among them mean, median and mode are called simple averages and the other two averages geometric mean and harmonic mean are called special averages.

The meaning of average is nicely given in the following definitions.

"A measure of central tendency is a typical value around which other figures congregate."

"An average stands for the whole group of which it forms a part yet represents the whole."

"One of the most widely used set of summary figures is known as measures of location."

**Characteristics for a good or an ideal average :**

The following properties should possess for an ideal average.

1. It should be rigidly defined.
2. It should be easy to understand and compute.
3. It should be based on all items in the data.
4. Its definition shall be in the form of a mathematical formula.
5. It should be capable of further algebraic treatment.
6. It should have sampling stability.
7. It should be capable of being used in further statistical computations or processing.

**Arithmetic mean or mean :**

Arithmetic mean or simply the mean of a variable is defined as the sum of the observations divided by the number of observations. If the variable x assumes n values $x_1, x_2 .. x_n$ then the mean, $\bar{x}$, is given by

$$\bar{x} = \frac{x_1 + x_2 + x_3 + .... + x_n}{n}$$

$$= \frac{1}{n}\sum_{i=1}^{n} x_i$$

This formula is for the ungrouped or raw data.

**Example 1 :**

Calculate the mean for 2, 4, 6, 8, 10

**Solution:**

$$\bar{x} = \frac{2+4+6+8+10}{5}$$

$$= \frac{30}{5} = 6$$

**Short-Cut method :**

Under this method an assumed or an arbitrary average (indicated by A) is used as the basis of calculation of deviations from individual values. The formula is

$$\bar{x} = A + \frac{\sum d}{n}$$

where, A = the assumed mean or any value in x

d = the deviation of each value from the assumed mean

## Harmonic mean (H.M) :

Harmonic mean of a set of observations is defined as the reciprocal of the arithmetic average of the reciprocal of the given values. If $x_1, x_2 \ldots x_n$ are n observations,

$$H.M = \frac{n}{\sum_{i=1}^{n}\left(\frac{1}{x_i}\right)}$$

For a frequency distribution

$$H.M = \frac{N}{\sum_{i=1}^{n} f\left(\frac{1}{x_i}\right)}$$

**Example 6:**

From the given data calculate H.M 5,10,17,24,30

| X | $\frac{1}{x}$ |
|---|---|
| 5 | 0.2000 |
| 10 | 0.1000 |
| 17 | 0.0588 |
| 24 | 0.0417 |
| 30 | 0.0333 |
| Total | 0.4338 |

$$H.M = \frac{n}{\Sigma\left[\frac{1}{x}\right]}$$

$$= \frac{5}{0.4338} = 11.526$$

0.2
5 ) 10
10
5

0.2
0.1
0.588
0.4167

$= \frac{50000}{4338}$

$= 11.52605$

**Median :**

The median is that value of the variate which divides the group into two equal parts, one part comprising all values greater, and the other, all values less than median.

**Ungrouped or Raw data :**

Arrange the given values in the increasing or decreasing order. If the number of values are odd, median is the middle value .If the number of values are even, median is the mean of middle two values.

By formula

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

**Example 11:**

When odd number of values are given. Find median for the following data

25, 18, 27, 10, 8, 30, 42, 20, 53

**Solution:**

Arranging the data in the increasing order 8, 10, 18, 20, 25, 27, 30, 42, 53

The middle value is the $5^{\text{th}}$ item i.e., 25 is the median

Using formula

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

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

$$= \left(\frac{10}{2}\right)^{\text{th}} \text{item}$$

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

$$= 25$$

**Mode :**

The mode refers to that value in a distribution, whic occur most frequently. It is an actual value, which has the highe concentration of items in and around it.

According to Croxton and Cowden " The mode of distribution is the value at the point around which the items tend be most heavily concentrated. It may be regarded at the m typical of a series of values".

It shows the centre of concentration of the frequency in aroun given value. Therefore, where the purpose is to know the poin the highest concentration it is preferred. It is, thus, a positi measure.

Its importance is very great in marketing studies whe manager is interested in knowing about the size, which has highest concentration of items. For example, in placing an order shoes or ready-made garments the modal size helps because sizes and other sizes around in common demand.

**Computation of the mode:**

**Ungrouped or Raw Data:**

For ungrouped data or a series of individual observations, mode is often found by mere inspection.

**Quartiles :**

The quartiles divide the distribution in four parts. There are three quartiles. The second quartile divides the distribution into two halves and therefore is the same as the median. The first (lower) quartile ($Q_1$) marks off the first one-fourth, the third (upper) quartile ($Q_3$) marks off the three-fourth.

**Raw or ungrouped data:**

First arrange the given data in the increasing order and use the formula for $Q_1$ and $Q_3$ then quartile deviation, Q.D is given by

$$Q.D = \frac{Q_3 - Q_1}{2}$$

Where $Q_1 = \left(\frac{n+1}{4}\right)^{th}$ item and $Q_3 = 3\left(\frac{n+1}{4}\right)^{th}$ item

**Example 22 :**

Compute quartiles for the data given below 25,18,30, 8, 15, 5, 10, 35, 40, 45

**Solution :**

5, 8, 10, 15, 18,25, 30,35,40, 45

**Deciles :**

These are the values, which divide the total number of observation into 10 equal parts. These are 9 deciles $D_1$, $D_2$…$D_9$. These are all called first decile, second decile. etc.,

**Percentiles :**

The percentile values divide the distribution into 100 parts each containing 1 percent of the cases. The percentile ($P_k$) is that value of the variable up to which lie exactly k% of the total number of observations.