# Measures of Central Tendency

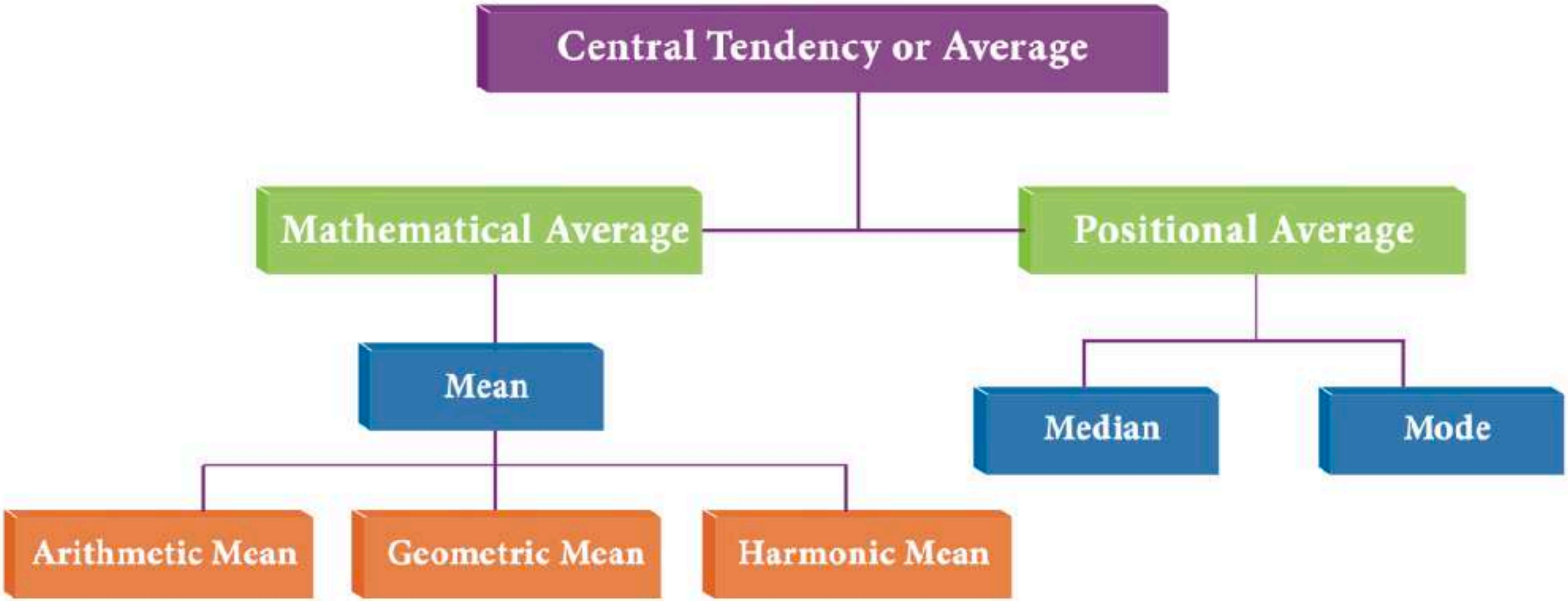
## Mean

Dr R Sinuvasan

# Measures of Central Tendency

The inherent inability of the human mind to grasp its entirely large body of numerical data compels us to see relatively few constants that will adequately describe the data - Prof. R. A. Fisher

Human mind is incapable of remembering the entire mass of unwieldy data. Having learnt the methods of collection and presentation of data, one has to condense the data to get representative numbers to study the characteristics of data. The characteristics of the data set is explored with some numerical measures namely measures of central tendency.


Central Tendency or Average
Mathematical Average
Positional Average
Mean
Median
Mode
Arithmetic Mean
Geometric Mean
Harmonic Mean

## Arithmetic Mean

**(a) To find A.M. for Raw data For a raw data, the arithmetic mean of a series of numbers is sum of all observations divided by the number of observations in the series.**

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

There are two methods for computing the A.M :

(i) Direct method (ii) Short cut method.

Example: The following data represent the number of books issued in a school library on selected from 7 different days 7, 9, 12, 15, 5, 4, 11 find the mean number of books.

$$\overline{x} = \frac{7+9+12+15+5+4+11}{7}$$

$$= \frac{63}{7} = 9$$

## (ii) Short-cut Method to find A.M.

Under this method an assumed mean or an arbitrary value (denoted by A) is used as the basis of calculation of deviations ($d_i$) from individual values. That is if $d_i = x_i - A$ then

$$\bar{x} = A + \frac{\sum_{i=1}^{n} d_i}{n}$$

Example: A student's marks in 5 subjects are 75, 68, 80, 92, 56. Find the average of his marks.

Let us take the assumed mean, A = 68

| $x_i$ | $d_i = x_i - 68$ |
|---|---|
| 75 | 7 |
| 68 | 0 |
| 80 | 12 |
| 56 | -12 |
| 92 | 24 |
| **Total** | **31** |

$$\bar{x} = A + \frac{\sum_{i=1}^{n} d_i}{n}$$

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

$$= 68+6.2 = 74.2$$

The arithmetic mean of average marks is 74.2

### (b) To find A.M. for Discrete Grouped data

If $x_1, x_2, \ldots, x_n$ are discrete values with the corresponding frequencies $f_1, f_2, \ldots, f_n$. Then the mean for discrete grouped data is defined as (direct method)

$$\bar{x} = \frac{\sum_{i=1}^{n} f_i x_i}{N}$$

In the short cut method the formula is modified as

$$\bar{x} = A + \frac{\sum_{i=1}^{n} f_i d_i}{N} \quad \text{where} \quad d_i = x_i - A$$

## Example

A proof reads through 73 pages manuscript The number of mistakes found on each of the pages are summarized in the table below Determine the mean number of mistakes found per page

| No of mistakes | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| No of pages | 5 | 9 | 12 | 17 | 14 | 10 | 6 |

(i) Direct Method

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

$$\bar{x} = \frac{\sum_{i=l}^{n} f_i x_i}{N}$$

$$= \frac{299}{73}$$

$$= 4.09$$

The mean number of mistakes is 4.09

(ii) Short-cut Method

| $x_i$ | $f_i$ | $d_i = x_i - A$ | $f_i d_i$ |
|---|---|---|---|
| 1 | 5 | -3 | -15 |
| 2 | 9 | -2 | -18 |
| 3 | 12 | -1 | -12 |
| 4 | 17 | 0 | 0 |
| 5 | 14 | 1 | 14 |
| 6 | 10 | 2 | 20 |
| 7 | 6 | 3 | 18 |
| | $\Sigma f_i = 73$ | | $\Sigma f_i d_i = 7$ |

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

$$= 4 + \frac{7}{73}$$

$$= 4.09$$

The mean number of mistakes $= 4.09$

### (c) Mean for Continuous Grouped data:

For the computation of A.M for the continuous grouped data, we can use direct method or short cut method.

#### (i) Direct Method:

The formula is

$$\overline{x} = \frac{\sum_{i=l}^{n} f_i x_i}{N}, \quad x_i \text{ is the midpoint of the class interval}$$

#### (ii) Short cut method

$$\overline{x} = \text{A} + \frac{\sum_{i=l}^{n} f_i d_l}{N} \times C$$

$$d = \frac{x_i - A}{c}$$

where A - any arbitrary value

c - width of the class interval

$x_i$ is the midpoint of the class interval.

The following the distribution of persons according to different income groups

| Income (in ₹1000) | 0 – 8 | 8 – 16 | 16 – 24 | 24 – 32 | 32 – 40 | 40 – 48 |
|---|---|---|---|---|---|---|
| **No of persons** | 8 | 7 | 16 | 24 | 15 | 7 |

Find the average income of the persons.

| *Class* | $f_i$ | $x_i$ | $f_i x_i$ |
|---|---|---|---|
| 0-8 | 8 | 4 | 32 |
| 8 – 16 | 7 | 12 | 84 |
| 16-24 | 16 | 20 | 320 |
| 24-32 | 24 | 28 | 672 |
| 32-40 | 15 | 36 | 540 |
| 40-48 | 7 | 44 | 308 |
| Total | N =77 | | 1956 |

$$\overline{x} = \frac{\sum_{i=l}^{n} f_i x_i}{N}$$

$$= \frac{1956}{77}$$

$$=25.40$$

## Short cut method:

| $Class$ | $f_i$ | $x_i$ | | $d_i$=( $x_i$–A)/c | $f_i d_i$ |
|---|---|---|---|---|---|
| 0 – 8 | 8 | 4 | | -3 | -24 |
| 8 – 16 | 7 | 12 | | -2 | -14 |
| 16 – 24 | 16 | 20 | | -1 | -16 |
| 24 – 32 | 24 | 28 | A | 0 | 0 |
| 32 – 40 | 15 | 36 | | 1 | 15 |
| 40 – 48 | 7 | 44 | | 2 | 14 |
| **Total** | **N= 77** | | | | **-25** |

$$\overline{x} = A + \frac{\sum_{i=1}^{n} f_i d_i}{N} \times C$$

$$= 28 + \frac{-25}{77} \times 8 \quad = 25.40$$

## Merits

- It is easy to compute and has a unique value.
- It is based on all the observations.
- It is well defined.
- It is least affected by sampling fluctuations.
- It can be used for further statistical analysis.

## Limitations

- The mean is unduly affected by the extreme items (outliers).
- It cannot be determined for the qualitative data such as beauty, honesty etc.
- It cannot be located by observations on the graphic method.

Arithmetic mean is a best representative of the data if the data set is homogeneous. On the other hand if the data set is heterogeneous the result may be misleading and may not represent the data.

# Weighted Arithmetic Mean

**Definition**

Let $x_1, x_2, ...., x_n$ be the set of n values having weights $w_1, w_2, ...., w_n$ respectively, then the weighted mean is,

$$\overline{x_w} = \frac{w_1 x_1 + w_2 x_2 + ... + w_n x_n}{w_1 + w_2 + ... + w_n} = \frac{\sum_{i=l}^{n} w_i x_i}{\sum_{i=l}^{n} w_i}$$

## Example

The weights assigned to different components in an examination or Component Weightage Marks scored

| Component | Weightage | Marks scored |
|---|---|---|
| Theory | 4 | 60 |
| Practical | 3 | 80 |
| Assignment | 1 | 90 |
| Project | 2 | 75 |
| | 10 | |

Calculate the weighted average score of the student who scored marks as given in the table

| Component | Marks scored ($x_i$) | Weightage ($w_i$) | $w_i x_i$ |
|---|---|---|---|
| Theory | 60 | 4 | 240 |
| Practical | 80 | 3 | 240 |
| Assignment | 90 | 1 | 90 |
| Project | 75 | 2 | 150 |
| Total | | 10 | 720 |

Weighted average, $\bar{x}_w = \dfrac{\sum w_i x_i}{\sum w_i}$

$= 720/10$

$= 72$

## Combined Mean:

Let $\overline{x}_1$ and $\overline{x}_2$ are the arithmetic mean of two groups (having the same unit of measurement of a variable), based on n1 and n2 observations respectively. Then the combined mean can be calculated using

$$\text{Combined Mean} = \bar{x}_{12} = \frac{n_1 \overline{x1} + n_2 \overline{x2}}{n_1 + n_2}$$

## Example

A class consists of 4 boys and 3 girls. The average marks obtained by the boys and girls are 20 and 30 respectively. Find the class average.

### *Solution:*

$n_1 = 4,\ \overline{x}_1 = 20,\ n_2 = 3,\ \overline{x}_2 = 30$

$$\text{Combined Mean} = \bar{x}_{12} = \frac{n_1 \overline{x1} + n_2 \overline{x2}}{n_1 + n_2}$$

$$= \left[\frac{4 \times 20 + 3 \times 30}{4 + 3}\right]$$

$$= \left[\frac{80 + 90}{7} = \frac{170}{7}\right] = 24.3$$

# Geometric Mean(GM)

## (a) G.M. For Ungrouped data

The Geometric Mean (G.M.) of a set of n observations is the nth root of their product. If $x_1, x_2, \dots, x_n$ are $n$ observations then

$$\text{G. M.} = \sqrt[n]{x_1, x_2 \dots x_n} = (x_1 . x_2 \dots x_n)^{\frac{1}{n}}$$

Taking the nth root of a number is difficult. Thus, the computation is done as under

$$\log \text{G.M.} = \log (x_1 . x_2 . . . x_n)$$

$$= (\log x_1 + \log x_2 + . . . + \log x_n)$$

$$= \frac{\sum_{i=l}^{n} \log x_i}{n}$$

$$\text{G.M.} = \text{Antilog} \frac{\sum_{i=l}^{n} \log x_i}{n}$$

## Example

Calculate the geometric mean of the annual percentage growth rate of profits in business corporate from the year 2000 to 2005 is given below

50, 72, 54, 82, 93

***Solution:***

| $x_i$ | 50 | 72 | 54 | 82 | 93 | **Total** |
|---|---|---|---|---|---|---|
| **log** $x_i$ | 1.6990 | 1.8573 | 1.7324 | 1.9138 | 1.9685 | **9.1710** |

$$\text{G.M.} = \text{Antilog} \frac{\sum_{i=1}^{n} \log x_i}{n}$$

$$= \text{Antilog} \frac{9.1710}{5}$$

$$= \text{Antilog } 1.8342$$

$$\text{G. M.} = 68.26$$

Geometrical mean of annual percentage growth rate of profits is 68.26

## (b) G.M. For Discrete grouped data

If $x_1, x_2, \ldots\ldots x_n$ are discrete values of the variate x with corresponding frequencies $f_1, f_2, \ldots f_n$. Then geometric mean is defined as

$$G.\,M. = \text{Antilog}\ \frac{\sum_{i=1}^{n} f_i \log x_i}{N} \text{ with usual notations}$$

**Example**

Find the G.M for the following data, which gives the defective screws obtained in a factory.

| Diameter (cm) | 5 | 15 | 25 | 35 |
|---|---|---|---|---|
| Number of defective screws | 5 | 8 | 3 | 4 |

| $x_i$ | $f_i$ | $\log x_i$ | $f_i \log x_i$ |
|---|---|---|---|
| 5 | 5 | 0.6990 | 3.4950 |
| 15 | 8 | 1.1761 | 9.4088 |
| 25 | 3 | 1.3979 | 4.1937 |
| 35 | 4 | 1.5441 | 6.1764 |
| | N=20 | | 23.2739 |

$$\text{G.M} = \text{Antilog}$$

$$= \text{Antilog} \; \frac{\sum_{i=l}^{n} f_i \log x_i}{N}$$

$$= \text{Antilog} \; \frac{23.2739}{20}$$

$$= \text{Antilog } 1.1637$$

$$\text{G.M} = 14.58$$

### (c) G.M. for Continuous grouped data

Let $x_i$ be the mid point of the class interval

$$\text{G. M.} = \text{Antilog}\left[\frac{\sum_{i=1}^{n} f_i \log x_i}{N}\right]$$

## Example

The following is the distribution of marks obtained by 109 students in a subject in an institution. Find the Geometric mean.

| Marks | 4-8 | 8-12 | 12-16 | 16-20 | 20-24 | 24-28 | 28-32 | 32-36 | 36-40 |
|---|---|---|---|---|---|---|---|---|---|
| No. of Students | 6 | 10 | 18 | 30 | 15 | 12 | 10 | 6 | 2 |

| Marks | Mid point ($x_i$) | $f_i$ | $\log x_i$ | $f_i \log x_i$ |
|---|---|---|---|---|
| 4-8 | 6 | 6 | 0.7782 | 4.6692 |
| 8-12 | 10 | 10 | 1.0000 | 10.0000 |
| 12-16 | 14 | 18 | 1.1461 | 20.6298 |
| 16-20 | 18 | 30 | 1.2553 | 37.6590 |
| 20-24 | 22 | 15 | 1.3424 | 20.1360 |
| 24-28 | 26 | 12 | 1.4150 | 16.980 |
| 28-32 | 30 | 10 | 1.4771 | 14.7710 |
| 32-36 | 34 | 6 | 1.5315 | 9.1890 |
| 36-40 | 38 | 2 | 1.5798 | 3.1596 |
| **Total** | | **N =109** | | **137.1936** |

$$\text{G.M.} = \text{Antilog}\left[\frac{\sum_{i=l}^{n} f_i \log x_i}{N}\right]$$

$$= \text{Antilog}\left[\frac{137.1936}{109}\right] = \text{Antilog}\,[1.2587]$$

G. M. = 18.14

**Merits of Geometric Mean:**

- It is based on all the observations
- It is rigidly defined
- It is capable of further algebraic treatment
- It is less affected by the extreme values
- It is suitable for averaging ratios, percentages and rates.

**Limitations of Geometric Mean:**

- It is difficult to understand
- The geometric mean cannot be computed if any item in the series is negative or zero.
- The GM may not be the actual value of the series
- It brings out the property of the ratio of the change and not the absolute difference of change as the case in arithmetic mean.

# Harmonic Mean (H.M.)

Harmonic Mean is defined as the reciprocal of the arithmetic mean of reciprocals of the observations.

## (a) H.M. for Ungrouped data

Let $x_1, x_2, \ldots, x_n$ be the $n$ observations then the harmonic mean is defined as

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

## Example

A man travels from Jaipur to Agra by a car and takes 4 hours to cover the whole distance. In the first hour he travels at a speed of 50 km/hr, in the second hour his speed is 65 km/hr, in third hour his speed is 80 km/hr and in the fourth hour he travels at the speed of 55 km/hr. Find the average speed of the motorist.

| $x$ | 50 | 65 | 80 | 55 | **Total** |
|---|---|---|---|---|---|
| $1/x$ | 0.0200 | 0.0154 | 0.0125 | 0.0182 | **0.0661** |

$$\text{H. M.} = \frac{n}{\sum\left(\frac{1}{x_i}\right)}$$

$$= \frac{4}{0.0661} = 60.5 \text{ km/hr}$$

Average speed of the motorist is 60.5km/hr

### (b) H.M. for Discrete Grouped data:

For a frequency distribution

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

## Example

The following data is obtained from the survey. Compute H.M

| Speed of the car | 130 | 135 | 140 | 145 | 150 |
|---|---|---|---|---|---|
| No of cars | 3 | 4 | 8 | 9 | 2 |

| $x_i$ | $f_i$ | $\frac{f_i}{x_i}$ |
|---|---|---|
| 130 | 3 | 0.0296 |
| 135 | 4 | 0.0091 |
| 140 | 8 | 0.0571 |
| 145 | 9 | 0.0621 |
| 150 | 2 | 0.0133 |
| Total | N = 26 | 0.1852 |

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

$$= \frac{26}{0.1852}$$

$$\text{H.M} = 140.39$$

### (c) H.M. for Continuous data:

The Harmonic mean $\text{H.M.} = \dfrac{N}{\sum_{i=1}^{n} f_i\left(\dfrac{1}{x_i}\right)}$

Where $x_i$ is the mid-point of the class interval

## Example

Find the harmonic mean of the following distribution of data

| Dividend yield (percent) | 2 – 6 | 6 – 10 | 10 – 14 |
|---|---|---|---|
| No. of companies | 10 | 12 | 18 |

| Class Intervals | Mid-value ($x_i$) | No. of companies ($f_i$) | Reciprocal ($1/x_i$) | $f_i$ ($1/x_i$) |
|---|---|---|---|---|
| 2 – 6 | 4 | 10 | ¼ | 2.5 |
| 6 – 10 | 8 | 12 | 1/8 | 1.5 |
| 10 – 14 | 12 | 18 | 1/12 | 1.5 |
| Total | | N = 40 | | 5.5 |

The harmonic mean is $\text{H.M.} = \dfrac{N}{\sum_{i=l}^{n} f_i\left(\dfrac{1}{x_i}\right)} = \dfrac{40}{5.5} = 7.27$

## Merits of H.M:

- It is rigidly defined
- It is based on all the observations of the series
- It is suitable in case of series having wide dispersion
- It is suitable for further mathematical treatment
- It gives less weight to large items and more weight to small items

## Limitations of H.M:

- It is difficult to calculate and is not understandable
- All the values must be available for computation
- It is not popular due to its complex calculation.
- It is usually a value which does not exist in series

*Harmonic mean is used to calculate the average value when the values are expressed as value/unit. Since the speed is expressed as km/hour, harmonic mean is used for the calculation of average speed.*

**Relationship among the averages:**

In any distribution when the original items are different the A.M., G.M. and H.M would also differ and will be in the following order:

$$\textbf{A.M.} \geq \textbf{G.M} \geq \textbf{H.M}$$

The mean of 100 items are found to be 30. If at the time of calculation two items are wrongly taken as 32 and 12 instead of 23 and 11. Find the correct mean.

A cyclist covers his first three kms at an average speed of 8 kmph. Another two kms at 3 kmph and the last two kms at 2 kmph. Find the average speed for the entire journey.