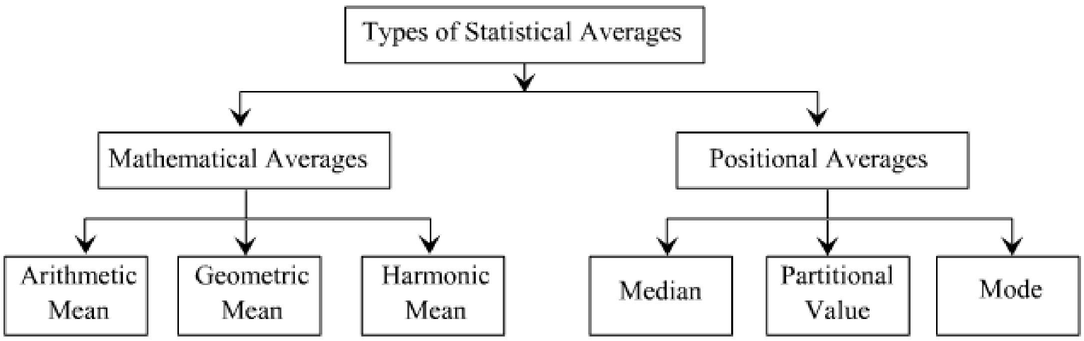
CHAPTER 5

MEASURES OF CENTRAL TENDENCY

- A **central tendency** refers to the central value or representative value of a statistical series.

- **Features of a Good Average**
  - Simple to calculate and easy to understand
  - Based on all values of observations
  - Capable of algebraic treatment
  - Least affected by extreme values
  - Least affected by fluctuation of sample

- **Types of Statistical Averages**



- **Arithmetic mean** is simply an average of all items in a series. It is donated by $\overline{X}$

For Individual Series		
Direct Method	$\overline{X} = \frac{X_1 + X_2 + X_3 + ...+X_n}{N}$ Or, $\overline{X} = \frac{\Sigma X}{N}$	where, $\overline{X}$: Arithmetic Mean $\sum X$: Sum of all observations N: Number of observations
Shortcut Method	$\overline{X} = A + \frac{\Sigma d}{N}$	where, A: Assumed mean Σd : Net sum of deviation of values from the assumed average N: Number of items in a series $d = X - A$
For Discrete Series		
Direct Method	$\overline{X} = \frac{\Sigma fX}{\Sigma f}$	where, ΣfX : Summation of the product of frequency and X values Σf : Sum total of frequency

Shortcut method	$\overline{X} = A + \frac{\Sigma fd}{\Sigma f}$	where, A: Assumed mean $d = X - A$ Σfd : Summation of the product of frequency and deviations Σf : Sum total of frequency
Step deviation method	$\overline{X} = A + \frac{\Sigma fd'}{\Sigma f} \times i$	where, $d' = \frac{X - A}{i}$ i = common factor
For Frequency Distribution Series (Continuous Series)		
Direct Method	$\overline{X} = \frac{\Sigma fm}{\Sigma f}$	where, Σfm : Sum of the product of frequency and mid values Σf : Sum total of frequency m : mid values $m = \frac{\text{Upper Limit} + \text{Lower Limit}}{2}$
Shortcut Method	$\overline{X} = A + \frac{\Sigma fd}{\Sigma f}$	where, $d = m - A$ A: Assumed mean $\Sigma f d$: Sum of the product of frequency and deviations Σf : Sum total of frequency
Step Deviation Method	$\overline{X} = A + \frac{\Sigma fd'}{\Sigma f} \times i$	where, $\Sigma f d'$ = Sum of the product of frequency and condensed deviations $d' = \frac{m - A}{i}$

Some Important Formulae

Corrected Arithmetic Mean	$\overline{X} = \frac{\Sigma X \text{ (Wrong Total)} + \text{Correct Value} - \text{Incorrect Value}}{N}$	
Weighted Arithmetic Mean	$\overline{X}_w = \frac{\Sigma WX}{\Sigma W}$	where, ΣWX : Summation of the product of weights and X values ΣW : Total weight

Combined Arithmetic Mean	$$\overline{X_{1,2}} = \frac{\overline{X}_1 N_1 + \overline{X}_2 N_2}{N_1 + N_2}$$	where, $\overline{X_{1,2}}$: Combined arithmetic mean of parts 1,2 of series $\overline{X_1}$: Arithmetic mean of part 1 of series $\overline{X_2}$: Arithmetic mean of part 2 of series N_1: Number of items of series 1 N_2: Number of items of series 2

- **Merits of Mean**
  - It is easy to calculate and understand
  - It is based on each and every observation.
  - It gives equal importance to all the items of a series.
  - It is not positional like median and mode, in fact, it is a calculated value.
  - It is free from any personal biasness.
  - It can be used for mathematical calculations, such as, multiplication, division, addition, etc.
  - It acts as a good base for comparison between two or more series.
  - It is stable measure of central tendency. That is, it is minimally affected by the change in the series.

- **Demerits of Mean**
  - Arithmetic mean is the most affected by the presence of extreme items. That is, it is easily distorted by the extreme values.
  - It cannot be calculated for an open-ended series.
  - It cannot be ascertained graphically.
  - It can be sometimes misleading and absurd results.
  - Sometimes, the computed value of the mean may not be from the given series.

- **Median** is the value that divides a series into two equal parts. It is a centrally located value of the series.

For Individual Series		
Steps	• Arrange different items of the series either in ascending or descending order. • Count the number of items in the series and denote it by N.	
	If N is an ***odd*** number $M = \text{Size of} \left(\frac{N+1}{2}\right)\text{th item}$	If N is an ***even*** number $M = \frac{\text{Size of}\left(\frac{N}{2}\right)\text{th item} + \text{Size of}\left(\frac{N}{2}+1\right)\text{th item}}{2}$

For Discrete Series	
Steps	• Arrange the data into ascending or descending order • Convert the simple frequency into cumulative frequency. • Sum the simple frequency $\sum f$ as N. • Determine the value of $\left(\frac{N+1}{2}\right)$th item • The value of X corresponding to $\left(\frac{N+1}{2}\right)$th item is the median value.
For Frequency Distribution (Continuous) Series	
Steps	• Convert the simple frequency into cumulative frequency. • Calculate the sum of simple frequencies i.e. $\sum f$ as N. • Determine the value of $\left(\frac{N}{2}\right)$th item • Median class corresponds to that cumulative frequency which includes the $\left(\frac{N}{2}\right)$th item • Compute median by the formula, $M = l_1 + \frac{\frac{N}{2} - CF}{f} \times i$ where, l_1 : lower limit of median class CF: Cumulative frequency of the class preceding the median class f: frequency of median class i: size of median class interval

- The value that divides the series into more than two parts is called **Partition Value**.

- A statistical series is divided into four equal parts; the end value of each part is called **quartile**.
  - The **first or lower quartile** is denoted by Q_1
  - The **second quartile** is same as median, i.e. Q_2 = Median
  - The **third quartile or upper quartile** is denoted by Q_3.

- **Calculation of Q_1 and Q_3**

Formulas to Calculate Quartile	
Lower Quartile (Q_1) **(Denotes top 25 % of the minimum values of a series)**	**Upper Quartile (Q_3)** **(Denotes lowest 25 % of the maximum values of a series)**
For Individual Series	
$Q_1 = \text{Size of } \left(\frac{N+1}{4}\right)^{\text{th}} \text{ item}$	$Q_3 = \text{Size of } \left(\frac{3(N+1)}{4}\right)^{\text{th}} \text{ item}$
For Discrete Series	
Locate the Size of $\left(\frac{N+1}{4}\right)^{\text{th}}$ item in the *CF* column and corresponding x value is Q_1	Locate the Size of $\left(\frac{3(N+1)}{4}\right)^{\text{th}}$ item in the *CF* column and corresponding x value is Q_3
For Continuous Series	
Locate the Size of $\left(\frac{N}{4}\right)^{\text{th}}$ item in *CF* column and the value of Q_1 will lie in the corresponding class interval. $Q_1 = l_1 + \frac{\frac{N}{4} - CF}{f} \times i$ where, l_1= Lower limit of class interval N = Sum of frequencies C*F* = Cumulative frequency of the class preceding the Q_1 class i = Class interval	Locate the Size of $\left(\frac{3N}{4}\right)^{\text{th}}$ item in *CF* column and the value of Q_3 will lie in the corresponding class interval. $Q_3 = l_1 + \frac{3\left(\frac{N}{4}\right) - CF}{f} \times i$ where, l_1= Lower limit of class interval N = Sum of frequencies C*F* = Cumulative frequency of the class preceding the Q_3 class i = Class interval

- The value that distributes the series into ten equal parts is called **Deciles** and is denoted by ***D***.

- **Calculation of Deciles**

Formulas to Calculate Deciles	
First Deciles (D_1)	**Ninth Deciles (D_9)**
For Individual Series	
$D_1 = \text{Size of } \left(\frac{N+1}{10}\right)^{\text{th}} \text{ item}$	$D_9 = \text{Size of } \left(\frac{9(N+1)}{10}\right)^{\text{th}} \text{ item}$
For Discrete Series	
Locate the Size of $\left(\frac{N+1}{10}\right)^{\text{th}}$ item in the *CF* column and corresponding *x* value is D_1	Locate the Size of $\left(\frac{9(N+1)}{10}\right)^{\text{th}}$ item in the *CF* column and corresponding *x* value is D_9
For Continuous Series	
Locate the Size of $\left(\frac{N}{10}\right)^{\text{th}}$ item in *CF* column and the value of D_1 will lie in the corresponding class interval. $D_1 = l_1 + \left(\frac{\frac{N}{10} - CF}{f}\right) \times i$ where, l_1 = Lower limit of class interval; N = Sum of frequencies; CF = Cumulative frequency of the class preceding the D_1 class; i = Class interval	Locate the Size of $\left(\frac{9N}{10}\right)^{\text{th}}$ item in *CF* column and the value of D_9 will lie in the corresponding class interval. $D_9 = l_1 + \left(\frac{9\left(\frac{N}{10}\right) - CF}{f}\right) \times i$ where, l_1 = Lower limit of class interval; N = Sum of frequencies; CF = Cumulative frequency of the class preceding the D_9 class; i = Class interval

- **Percentile** is defined as that value of a series that divides the series into 100 equal parts. It is denoted by ***P***.

Calculation of Percentiles

Formulas to Calculate Percentiles	
First Percentiles (P_1)	**Ninth Percentiles (P_{90})**
For Individual Series	
$P_1 = \text{Size of } \left(\frac{N+1}{100}\right)^{th} \text{ item}$	$P_{90} = \text{Size of } \left(\frac{90(N+1)}{100}\right)^{th} \text{ item}$
For Discrete Series	
Locate the Size of $\left(\frac{N+1}{100}\right)^{th}$ item in the *CF* column and corresponding *x* value is P_1	Locate the Size of $\left(\frac{90(N+1)}{100}\right)^{th}$ item in the *CF* column and corresponding *x* value is P_{90}
For Continuous Series	
Locate the Size of $\left(\frac{N}{100}\right)^{th}$ item in *CF* column and the value of P_1 will lie in the corresponding class interval. $P_1 = l_1 + \left(\frac{\frac{N}{100} - CF}{f}\right) \times i$ where, l_1= Lower limit of class interval; N = Sum of frequencies; CF = Cumulative frequency of the class preceding the P_1 class; i = Class interval	Locate the Size of $\left(\frac{90N}{100}\right)^{th}$ item in *CF* column and the value of P_{90} will lie in the corresponding class interval. $P_{90} = l_1 + \left(\frac{90\left(\frac{N}{100}\right) - CF}{f}\right) \times i$ where, l_1= Lower limit of class interval; N = Sum of frequencies; CF = Cumulative frequency of the class preceding the P_{90} class; i = Class interval

Merits of Median

- It is easy to calculate and understand
- Unlike mean, median is not affected by the presence of the extreme values.
- Unlike mean, median can be graphically ascertained.
- Unlike mean, the computed value of median is always reflected in the series.
- It is most appropriate for the open-ended series.
- It is a positional value and not a calculated value like, mean.

Demerits of Median

- As it is a positional average, so it is not based on all the observations of a series.
- It requires observations to be arranged in either ascending or descending order.
- It does not have a precise value (instead is an approximate value) when the number of observations is even (in case of individual series).

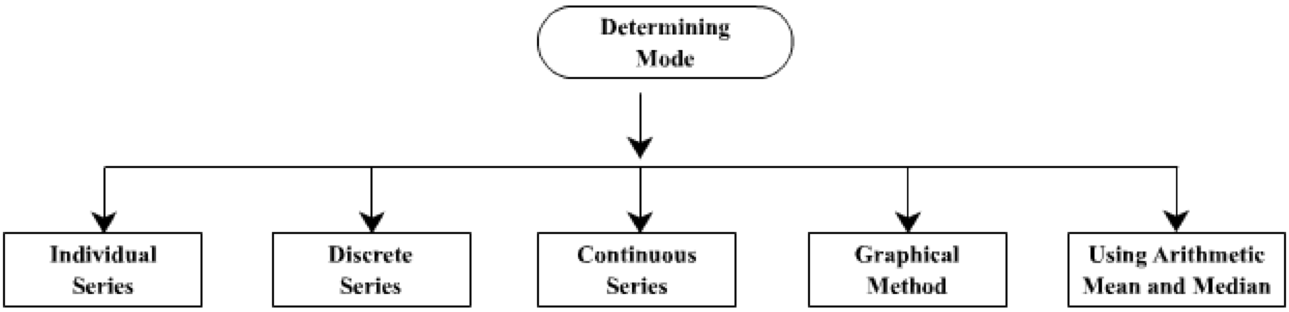
- Unlike mean, median is not capable of further mathematical treatment, such as, addition, subtraction, multiplication, etc.

- **Mode** is defined as the value that occurs (or repeats itself) most frequently in a series. It is denoted by *Z*.

- **Merits of Mode**
  - It is unaffected by the presence of large or small value in the series.
  - It is unaffected by the presence of extreme values.
  - It indicates the value of the series in the best way.
  - It is comparatively easier to calculate mode, as the information regarding all values and all frequencies is not required. Only the most frequent (maximum) value is to be known.

- **Limitations of Mode**
  - It is not based on all values of a series.
  - It is incapable of any further algebraic treatment.
  - Sometimes, it is difficult to ascertain the definite value of mode.
  - It fails to represent the small values of a series, therefore, may not be the best indicator of the series.

- **Calculation of Mode**



For Individual Series	
Inspection Method	The most frequent value in a series is identified as mode.
By changing the series into discrete series	If the number of items in the series is very large, then the value corresponding to the highest frequency is identified as mode.
For Discrete Series	
Inspection Method	The value that has the highest frequency in the series is identified as mode.
Grouping Method	A grouping table and analysis table is prepared under this method.

For Frequency Distribution (Continuous) Series		
Inspection Method	The modal class that has the maximum frequency is to be identified and the modal value is calculated by the following formula: $Z = l_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i$	where, Z = value of Mode; l_1 = lower limit of modal class; f_0 = Frequency of the preceding modal class; f_1 = Frequency of the modal class; f_2 = Frequency of the subsequent modal class or post modal class; i = Class interval of the modal class.
Grouping Method	The modal class is obtained from the grouping table and the analysis table. The modal value is calculated as: $Z = l_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i$	where, Z = value of Mode; l_1 = lower limit of modal class; f_0 = Frequency of the preceding modal class; f_1 = Frequency of the modal class; f_2 = Frequency of the subsequent modal class or post modal class; i = Class interval of the modal class.

Graphical Method of Ascertaining Mode

- Draw histograms (rectangles) for different class intervals.
- Identify the highest rectangle. This gives the modal class interval.
- The top right corner of the highest histogram is joined with the top right corner of the immediate left histogram.
- The top left corner of the highest histogram is joined with the top left corner of the immediate right histogram.
- The intersection point of the two lines is extended to the *x*-axis by drawing a perpendicular. The point on the x-axis is the modal value of the series.

Important Notes for Calculating Mode

- *If mode is to be calculated for inclusive series, then the inclusive series is to be converted into exclusive series by deducting 0.5 from the lower limit and adding 0.5 to the upper limit of each class intervals.*
- *If mid-values are given instead of the class intervals, then for calculating mode, we firs, need to convert the mid-values into class intervals (i.e. into continuous series), then mode is calculated as is done in the case of continuous series.*

Relation between Mean, Median and Mode

- ***Empirical Relationship***

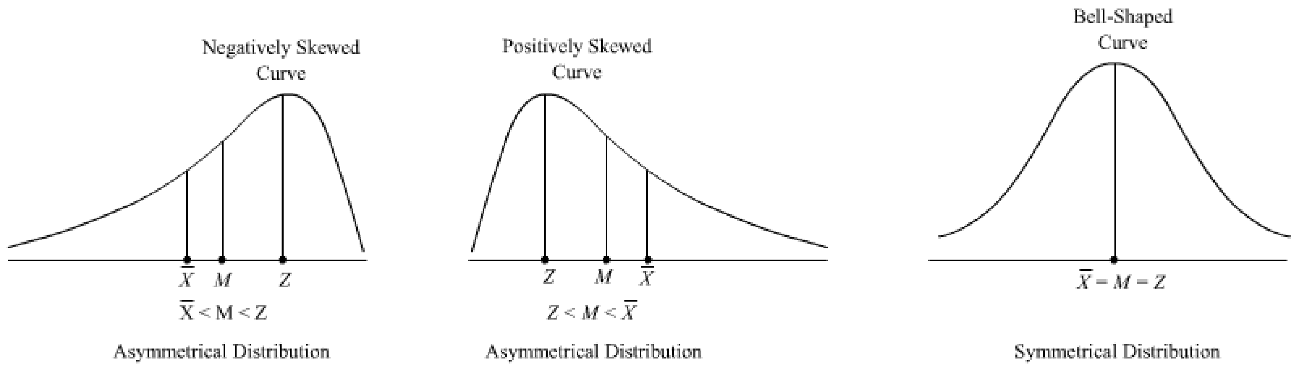
  $Z = 3M - 2\overline{X}$ (For all types of distribution- Symmetrical and Asymmetrical)

  where,

  Z= Mode

  M = Median

  $\overline{X}$ = Mean



- $\overline{X} < M < Z$ (For Negatively Skewed Distribution)
- $\overline{X} = M = Z$ (For Symmetrical Distribution)
- $\overline{X} > M > Z$ (For Positively Skewed Distribution)

Relationship between Percentiles, Deciles, Quartiles and Median

- $P_{50} = D_5 = Q_2$ = Median
- $P_{10} = D_1$
- $P_{20} = D_2$