Page No.			
Date			

2. Measures of Central Tendency

* Average → central value

Average
- Mathematical Average
  - Arithmetic mean
  - Geometric mean
  - Harmonic mean
- Positional Average
  - median
  - Quartiles
  - Deciles
  - Percentaile

  } Based on position
  - Mode → Based on Repetition

	Avg useful
* Homogeneous data (Equal values)	No
Heterogeneous data (unequal /distint values)	Yes

* Arithmetic mean $(\bar{X})$

→ Sum of all observation divided by no. of observation.

→ $\bar{X} = \frac{\Sigma x}{n}$ or $\frac{\Sigma fx}{\Sigma f}$

→ Best Avg. as compared to median & mode

→ Rigidly defined as compared to median & mode.

→ Affected by extreme values.

Date

* Shortcuts:

1] AM of 1st 'n' natural no → $\frac{n+1}{2}$

1, 2, 3, 4, 5, 6, 7

$\bar{x} = \frac{7+1}{2} = 4$

2] AM of 1st 'n' odd natural no → n

Eg: 1, 3, 5, 7, 9, 11

AM = 6

3] AM of 1st 'n' even natural no → n+1

Eg. 2, 4, 6, 8, 10

AM = 5+1 = 6.

4] AM of equidistant values

$= \frac{1^{st}\text{ value} + \text{last value}}{2}$

1^{st} value → Smallest

Last value → largest

Note:- Ascending order required

* Combined AM

→ $\bar{X}_c = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2}$

→ Avg. of all averages.

Page No.	
Date	

* Weighted AM

→ weights considered as frequency
→ It is used when all values are not equal importance.
→ weighted AM of 1st 'n' natural no. when values of X are equal to corresponding weights.

$= \frac{2n+1}{3}$ ←

x	w		x	f
1	1		1	1
2	2		2	2
3	3	OR	3	3
4	4		4	4
5	5		5	5
6	6		6	6

→ $\bar{X}_w = \frac{\Sigma wx}{\Sigma w}$

where w = weight.

* Geometric Mean

→ n^{th} root of product of all observation.
→ If any value is '0' or '-ve' then GM cannot be determined.
→ Mainly used to fined Avg. rates (or) ratio's of quantities which are changing at a cummulative rates.
→ It is less affected by sampling fluctuation compared to mode & median.

Page No

Date

→ It is less affected by extreme value compared to AM.

→ GM is extensively used to construct index no.

→ $GM = \sqrt[n]{x_1 \cdot x_2 \cdot x_3 \cdots x_n}$

(or)

$\sqrt[N]{(x_1)^{f_1} \cdot (x_2)^{f_2} \cdots (x_n)^{f_n}}$

where $N = \Sigma f$.

→ If $z = xy$

then GM of z = (GM of x)(GM of y)

→ If $z = x/y$

then GM of z = $\frac{\text{GM of } x}{\text{GM of } y}$

→ GM is used when change is related to immediate preceeding data.

E.g Avg. rate of depreciation

(or)

Avg. rate of growth of population

(OR)

Avg. ROI (compound interest)

→ Combined GM $= \sqrt[n_1+n_2]{(G_1)^{n_1}(G_2)^{n_2}}$

Page No.
Date

* How to find Roots on calc.

- Take desired ~~root~~ no. & press $\boxed{\sqrt{\ }}$ 12 times.
- $-1 =$
- $\div \boxed{\text{desired root}} =$
- $+1 =$
- Press
  $\times =$
  $\times =$
  $\times =$
  ⋮ 12 times.

desired root

$\sqrt[3]{8} = 2.$

Desired no.

1, 2, 3 → $1, \frac{1}{2}, \frac{1}{3}$

AM = $\boxed{2}$ → $HM = \frac{1}{2}$

$HM = \frac{3}{\frac{1}{1/1} + \frac{1}{1/2} + \frac{1}{1/3}}$

$= \frac{3}{1+2+3} = \frac{1}{2}.$

* Harmonic Mean (HM)

→ Reciprocal of AM → Note: Data Bhi reciprocal hona chahiye.

→ No. of observation divided by sum of reciprocal of all observations.

→ If any value is '0' then reciprocal of '0' is undefined, so HM cannot be calculated.

→ HM has a very restricted use & they are usually used for calculating avg. speed & avg. rates of quantities etc.

→ less affected by extreme values & sampling fluctuation compare to AM and GM.

→ $HM = \dfrac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \frac{1}{x_3} + \cdots + \frac{1}{x^n}}$

OR

$$\frac{\Sigma f}{\frac{f_1}{x_1} + \frac{f_2}{\cancel{f}x_2} + \frac{f_3}{x_3} + \cdots + \frac{f_n}{x_n}}$$

→ Combined $HM = \dfrac{n_1 + n_2}{\frac{n_1}{H_1} + \frac{n_2}{H_2}}$

* Relation betn AM, GM, HM :

1) Homogeneous data → $\boxed{AM = GM = HM}$

Heterogeneous data / unequal values → $\boxed{AM > GM > HM}$

Generally

$$\boxed{AM \geq GM \geq HM}$$

2) $GM = \sqrt{(AM)(HM)}$ $\qquad GM^2 = (AM)(HM)$

3) If only 2 observations → a & b

$AM = \dfrac{a+b}{2}$, $GM = \sqrt{ab}$, $HM = \dfrac{2ab}{a+b}$

* 2 unit concept

Km/hr

	Distance	Time
Given Constant		

If tick comes here then use HM

If tick comes here then use AM

* Median.

→ Middle most value of entire observation.

→ divides data into 2 equal parts.

→ Half of the data lies above median & half of the data lies below median.

→ Best avg. after AM.

→ Best avg for [open class interval]

means

[lower limit of class not given (or) upper limit of class not given (or) Both are not given]

→ Sum of absolute deviation taken from median always be minimum.

Page No.
Date

$$\boxed{\Sigma \left| x - \text{median} \right| = \text{minimum}}$$

* **Remember**

$\Sigma (x - \bar{x}) = 0$

Sum of deviation taken from AM always be zero.

* **Quartiles**

→ divides data into 4 equal parts.

→ Q_1 , Q_2 , Q_3

Q_1 ← 1st quartile or lower quartile

Q_2 → median or 2nd Quartile

Q_3 ← 3rd Quartile or upper quartile.

→ $\underbrace{25\% \mid 25\%}_{50\%} \mid \underbrace{25\% \mid 25\%}_{50\%}$

Q_1 , Q_2 , Q_3

* **Deciles**

→ Divides data into 10 equal parts

→ $\underbrace{10\% \mid 10\% \mid 10\% \mid 10\% \mid 10\%}_{50\%} \mid \underbrace{10\% \mid 10\% \mid 10\% \mid 10\% \mid 10\%}_{50\%}$

D_1 D_2 D_3 D_4 D_5 D_6 D_7 D_8 D_9

D_4 ← 4th decile

D_5 ← 5th decile or median.

* Percentiles

→ P_1 P_2 ---- P_{50} ---- P_{99}

↖ 50th percentile (or) Median.

$Q_1 = P_{25}$	$Q_3 = P_{75}$	
$Q_2 = D_5 = P_{50}$		
$D_1 = P_{10}$	$D_4 = P_{40}$	$D_7 = 70$
$D_2 = P_{20}$	$D_5 = P_{50}$	$D_8 = 80$
$D_3 = P_{30}$	$D_6 = P_{60}$	$D_9 = 90$

	Raw Data	Discrete series	Continuous series
	3, 5, 8, 7, 10, 9	X | f: 10 | 3, 20 | 8, 30 | 5	Class | f: 10–20 | 3, 20–30 | 8, 30–40 | 5
Quartile (Q_K)	$K\left(\frac{N+1}{4}\right)^{th}$ obs.	$K\left(\frac{N+1}{4}\right)^{th}$ obs.	$l_1 + \left(\frac{\frac{KN}{4} - cf}{f}\right) \times h$
Decile (D_K)	$K\left(\frac{N+1}{10}\right)^{th}$ obs.	$K\left(\frac{N+1}{10}\right)^{th}$ obs.	$l_1 + \left(\frac{\frac{KN}{10} - cf}{f}\right) \times h$
Percentile (P_K)	$K\left(\frac{N+1}{100}\right)^{th}$ obs.	$K\left(\frac{N+1}{100}\right)^{th}$ obs.	$l_1 + \left(\frac{\frac{KN}{100} - cf}{f}\right) \times h$
Median	$\left(\frac{N+1}{2}\right)^{th}$ obs.	$\left(\frac{N+1}{2}\right)^{th}$ obs.	$l_1 + \left(\frac{\frac{N}{2} - cf}{f}\right) \times h$

Page No.
Date

Notes	Arrange data into ascending order	1) X should be in ascending order 2) Find less than c.f.	1) Data should be continuous series 2) Find less than cf 3) $l_1 \rightarrow$ lower limit $l_2 \rightarrow$ Upper limit cf $\rightarrow$ Cumulative frequency $h \rightarrow$ class width $(l_2 - l_1)$ 4) To determine class always find first N value part $\frac{N}{2}, \frac{KN}{4}, \frac{KN}{10}, \frac{KN}{100}$

* Mode

→ Value repeats maximum no. of times.

→ Most unstable but quickest method to find avg.

→ It is not affected by extreme observation but most affected by sampling fluctuation.

→ It can be located graphically with the help of histogram (or) Area diagram (or) frequency diagram.

→ For raw data
Mode → Value repeats
$\boxed{\text{Zyada bar}}$

Page No.
Date

→ For discrete series
Mode → value from X (jiski frequency highest hai)

→ Continuous series

$$\text{Mode} = L_1 + \left(\frac{f_m - f_1}{2f_m - f_1 - f_2}\right) \times h$$

f_m → maximum frequency
f_1 → above f_m h → class width
f_2 → Below f_m $l_2 - l_1$

* Relation betn AM, Median & Mode.

1) Homogeneous data [Equal value]
(Symmetrical data)
↓
$\boxed{AM = Median = Mode}$

2) Asymmetrical data [Heterogeneous]

↓ Positively skewed — Mean > Median > Mode

↓ Negatively skewed — Mean < Median < mode

3) Mean − mode = 3 (mean − median)

OR

Mode = 3 median − 2mean