15) Relative frequency = $\frac{Class\ frequency}{Total\ frequency}$

16) Relative frequency lies between 0 to 1.

17) Percentage frequency = $\frac{Class\ frequency}{Total\ frequency} \times 100$

18) $Frequency\ density = \frac{Class\ frequency}{class\ width}$

2. MEASURES OF CENTRAL TENDENCY

PART - 1

INTRODUCTION

1) Average is a single expression which represents the whole group, and it gives adequate idea about the group.
2) There are 3 types of averages: Mean, Median & Mode
3) Mean is of three types: AM, GM & HM
4) AM, GM & HM are Mathematical averages. Median and Mode are positional averages.

ARITHMETIC MEAN

1) AM is the best measure of central tendency, and it is the mostly widely used measure to represent the entire data.
2) A.M. is the most stable, reliable and ideal measure of central tendency.
3) AM is defined as the sum of observations divided by number of observations.

INDIVIDUAL SERIES: $\bar{x} = \frac{\Sigma x}{n} \Rightarrow \Sigma x = n\bar{x}$ (n = number of values, $\bar{x}$ = arithmetic mean)

1) If the values are in A.P. then $\bar{x} = \frac{a+l}{2}$ ('a' is first term and l is the last term.)
2) Assumed mean method: $\bar{x} = A + \frac{\Sigma d}{n}$; Where d = x-A (Any random number can be chosen as Assumed Mean)

DISCRETE SERIES: $\bar{x} = \frac{\Sigma fx}{\Sigma f}$

CONTINUOUS SERIES: Same as discrete series. But x means mid values.

Step deviation method for continuous series, $\bar{x} = A + \frac{\Sigma fd}{\Sigma f} \times c$; (d = x - A, 'c' is the class length).

eg- 5, 5, 5
AM = 5

IMPORTANT PROPERTIES OF AM:

1) If all the observations assumed by a variable are constant, then AM is also same constant.
2) The choice of assumed mean does not affect the original mean.
3) The algebraic sum of deviations of a set of observations from their AM is zero.
4) The sum of squares of deviations taken from AM is always minimum.

diff square

5) AM is dependent on origin and scale i.e., arithmetic mean is affected due to change of origin and scale.
6) If 2 variables are related by linear equation Y = a + b X, then AM of Y is given by $\bar{Y} = a + b\bar{X}$
7) The mean of first n natural numbers is given as $\frac{n+1}{2}$
8) The mean of square of first n natural numbers is $\frac{(n+1)(2n+1)}{6}$
9) The mean of cubes of first n natural numbers is $\frac{n(n+1)^2}{4}$

COMBINED AM: Also called pooled mean.

$\bar{x}_{12} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1+n_2}$, For 3 groups

$\bar{x}_{123} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2 + n_3\bar{x}_3}{n_1+n_2+n_3}$

If $n_1 = n_2 = n_3$ then $\bar{x}_{123} = \frac{x_1 + x_2 + x_3}{3}$

If $n_1 = n_2 = n_3$ and $\bar{x}_1 = \bar{x}_2 = \bar{x}_3$ then $\bar{x}_{123} = \bar{x}_1$ or $\bar{x}_2$ or $\bar{x}_3$

CORRECTED AM.: $\bar{x}_c = \bar{x} + \frac{c - w}{n}$ (where c- correct value and w- wrong value)

NOTE: AM is rigidly defined, easy to comprehend, simple to calculate and possess mathematical properties. The main disadvantages of the mean are it is highly affected by extreme observations and not suitable for open end classification.

WEIGHTED AM:

AM gives equal importance to all. When it is needed to give unequal importance to items then we will use weighted AM. Weighted A.M.

$$\boxed{\bar{x}_w = \frac{\Sigma wx}{\Sigma w}}$$

$\bar{x}_w = \frac{\Sigma wx}{\Sigma w}$

uniformity in variability

1) This is similar to AM for discrete series (Frequency is replaced by Weights)

2) Weighted AM of first n natural numbers with corresponding weights is $\frac{2n+1}{3}$

MEDIAN

1) It is the middle-most or central value when the observations are arranged either in ascending order or in descending order of their magnitude. Also known as Positional average.

2) Best measure for qualitative data and suitable average in case of open-end distribution.

3) Can be obtained from Ogive curve also. (Other partition values like quartiles, deciles, percentiles can be obtained from ogive curve)

INDIVIDUAL SERIES: First arrange data in ascending or descending order.

$M = \left(\frac{n+1}{2}\right)^{th}$ term, if n is odd

M= Average of $\frac{n}{2}$ and $\frac{n}{2}$+1 terms, if n is even

The rank of median = $\frac{n+1}{2}$; if n is odd

= Average of $\frac{n}{2}$ and $\frac{n}{2}$ + 1; if n is even.

DISCRETE SERIES:

1) The data must be arranged in increasing order of magnitude. (x values)

2) Find the cumulative frequencies (L.C.F)

3) Formula will be similar to individual series except that n will be replaced by N, where N = sum of frequencies = Σf

CONTINUOUS SERIES: Must be exclusive. Find less than cumulative frequencies.

Median (M) $= L + \frac{\frac{N}{2}-m}{f} \times c$ (The class in which N/2 frequency lies is median class.)

L = lower class boundary of the median class (i.e., the class containing median.)

m= Preceding less than cumulative frequency value corresponding to N/2. (Pre median class)

f = frequency of median class, N = total frequency, C = length of median class

PROPERTIES:

1) The sum of the absolute deviations taking from their median is minimum.

2) Median is affected due to change of origin and scale.

If two variables X and Y are related by the linear equation Y = a+ b X, then the median of y is given by $M_y = a + b M_x$

PARTITION VALUES

	QUARTILES	DECILES	PERCENTILES
No of parts	4	10	100
No of values	3	9	99
Range	$Q_3 - Q_1$	$D_9 - D_1$	$P_{90} - P_{10}$
Median value	Q_2	D_5	P_{50}

PARTITION VALUE	INDIVIDUAL SERIES	DISCRETE SERIES	CONTINUOUS SERIES
QUARTILES	$Q_k = k\left(\frac{n+1}{4}\right)^{th}$ term Rank of $Q_k = k\left(\frac{n+1}{4}\right)$ where, k = 1, 2, 3	$Q_k = k\left(\frac{N+1}{4}\right)^{th}$ term Rank of $Q_k = k\left(\frac{N+1}{4}\right)$ k = 1, 2, 3	$Q_k = l + \frac{\frac{kN}{4} - cf}{f} \times h$ Rank of $Q_k = \frac{kN}{4}$ k = 1, 2, 3

DECILES	$D_k = k\left(\frac{n+1}{10}\right)^{th}$ term Rank of $D_k = k\left(\frac{n+1}{10}\right)$ where, k = 1, 2, 3, ... 9	$D_k = k\left(\frac{N+1}{10}\right)^{th}$ term Rank of $D_k = k\left(\frac{N+1}{10}\right)$ k = 1, 2, 3, ...9	$D_k = l + \frac{\frac{kN}{10} - cf}{f} \times h$ Rank of $D_k = \frac{kN}{10}$ k = 1, 2, 3, ... 9
PERCENTILES	$P_k = k\left(\frac{n+1}{100}\right)^{th}$ term Rank of $P_k = k\left(\frac{n+1}{100}\right)$ where, k = 1, 2, 3, ... 99	$P_k = k\left(\frac{N+1}{100}\right)^{th}$ term Rank of $P_k = k\left(\frac{N+1}{100}\right)$ k = 1, 2, 3, ...99	$P_k = l + \frac{\frac{kN}{100} - cf}{f} \times h$ Rank of $P_k = \frac{kN}{100}$ k = 1, 2, 3, ... 99

3. MEASURES OF CENTRAL TENDENCY

PART - 2

MODE

1) It is defined as the value that occurs maximum number of times
2) It is not an isolated value like mean or median and it may be uni -modal or bi-modal.
3) Sometimes it is ill-defined or ill-determinate.
4) It can be obtained from Histogram.
5) Mode is affected due to change of origin and scale. If two variables X and Y are related by the linear equation that Y = a + b X, then mode of y is given by $Z_y = a + bZ_x$
6) For an ungrouped or discrete frequency distributions mode is value of variable which has the maximum or highest frequency.

CONTINUOUS SERIES:

$$\text{Mode} = L + \left(\frac{f - f_1}{2f - f_1 - f_2}\right) \times c$$

$$= L + \left(\frac{f - f_1}{(f - f_1) + (f - f_2)}\right) \times c$$

$$= L + \left(\frac{\Delta_1}{\Delta_1 + \Delta_2}\right) \times c$$

$$= L + \left(\frac{d_1}{d_1 + d_2}\right) \times c$$

Where $\Delta_1 = d_1 = f - f_1$, $\Delta_2 = d_2 = f - f_2$ and

L = LCB of the modal class.

f = frequency of the modal class

f1 = frequency of the pre-modal class

f2 = frequency of the post modal class

C = class length of the modal class UB-LB

NOTE: The class with highest frequency is called modalclass.

FEATURES OF IDEAL MEASURE OF CENTRAL TENDENCY:

FEATURE OF AVERAGE	ARITHMETIC MEAN	MEDIAN	MODE
Rigidly defined	Yes	No	No
Easy to understand	Yes	Yes	Yes
No. of observations	It is based on all the observations	It is based on only middle most observations	It is based on the observations which are repeated mostly.
Effect of extreme observations	Highly effected	Not effected	Least effected
Effect of sampling fluctuations	Least effected	Highly effected	Highly effected

Graphical representation	Cannot be represented graphically	Ogive curve	Histogram
Mathematical treatment	Yes	No	No

OTHER POINTS TO REMEMBER:

	ARITHMETIC MEAN	MEDIAN	MODE
Use	Most widely used	Not so popular	Not so popular
Open end classes	Not suitable	Most Suitable	suitable
Qualitative characteristics	Cannot be determined	Most suitable	suitable
Inclusive series is given	No need to convert into exclusive	Need to convert in to exclusive	Need to convert in to exclusive
Origin and scale	Effected by both origin and scale.	Effected by both origin and scale	Effected by both origin and scale
If X and Y are linearly related by Y = a +b X	$\bar{y} = a + b\bar{x}$	$M_y = a + b M_x$	$Z_y = a + bZ_x$

RELATIONSHIP BETWEEN MEAN, MEDIAN AND MODE.

a)	Symmetrical Distribution	Mean = Median = Mode
b)	Positively Skewed Distribution	Mean > Median > Mode
c)	Negatively Skewed Distribution	Mean < Median < Mode
d)	Moderately Asymmetrical Distribution	Mode = 3 Median – 2 Mean, Mean – mode = 3 (mean – median) Median - mode = $\frac{2}{3}$ (mean – mode)

GEOMETRIC MEAN

INDIVIDUAL SERIES: GM is the n^{th} root of product of all items in a series

$$GM = \sqrt[n]{x_1.x_2.x_3..x_n}$$

Alternatively, GM = Antilog $\left(\frac{\Sigma \log x}{n}\right)$

For two values GM = $\sqrt{ab}$ and for three values GM = $\sqrt[3]{abc}$

PROPERTIES:

1) If one of observations is zero, then GM is Zero.
2) If all the observations are having equal value, then the same value is the GM of the values.
3) GM is used in the construction of index numbers.
4) GM (x y) = GM (x) × GM (y)
5) GM (x/y) = GM (x) ÷ GM (y)
6) $GM_{12} = \text{antilog}\left[\frac{n_1 \log GM_1 + n_2 \log GM_2}{n_1 + n_2}\right]$
7) Both GM and HM are called as ratio averages.
8) It is difficult to determine as it involves logarithms, roots, ratios etc.,

DISCRETE SERIES & CONTINUOUS SERIES:

$$GM = \sqrt[N]{x_1^{f_1}.x_2^{f_2}.x_3^{f_3}......x_n^{f_n}} \; ; \; N = \Sigma f$$

Alternatively, GM = Antilog $\left(\frac{\Sigma f.\log x}{\Sigma f}\right)$ (For continuous series we take x= mid value)

WEIGHTED GM:

Weighted GM = $\sqrt[\Sigma w]{x_1^{w_1} \cdot x_2^{w_2} \ldots x_n^{w_n}}$,

Alternatively, Weighted GM = Antilog $\left(\frac{\Sigma w.\log x}{\Sigma w}\right)$

HARMONIC MEAN:

1) HM is the reciprocal of AM of reciprocals of observations.
2) HM is useful in averaging rates, ratios, speed, and prices.

a) INDIVIDUAL SERIES:

$$HM = \frac{n}{\frac{1}{x_1}+\frac{1}{x_2}+\cdots\cdots+\frac{1}{x_n}} = \frac{n}{\Sigma\frac{1}{x}}$$

In case of two observations a and b,

$$HM = \frac{2ab}{a+b}$$

b) DISCRETE SERIES & CONTINUOUS SERIES:

$HM = \frac{\Sigma f}{\Sigma\frac{f}{x}}$ (For continuous series, x= mid value)

c) WEIGHTED HM = $\frac{\Sigma w}{\Sigma\frac{w}{x}}$

d) COMBINED H.M.:

For two groups, combined HM = $\frac{n_1+n_2}{\frac{n_1}{H_1}+\frac{n_2}{H_2}}$

i) For equal distances we use simple HM to determine average speed and for unequal distances we use weighted HM.

ii) For equal quantities we use simple AM to determine average price and for unequal quantities we use simple HM.

RELATIONSHIP BETWEEN AM, GM AND HM:

a) For equal observations	AM = GM = HM
b) For distinct observations	AM > GM > HM
c) For set of positive observations	AM ≥ GM ≥ HM
d) For a set of two observations	GM² = AM x HM