MEASURES OF CENTRAL TENDENCY & DISPERSION, SKEWNESS, KURTOSIS

Central Tendency and Dispersion are **the most common and widely used statistical tool** which handles large quantity of data and reduces the data to a single value used for doing comparative studies and draw conclusion with accuracy and clarity.

MAIN OBJECTIVES

1. To **condense data** in a single value.
2. To facilitate **comparisons** between data.

the tendency of data to cluster around a central or mid value is called central tendency of data, central tendency is measured by averages.

REQUISITES OF CENTRAL TENDENCY

a. It should be **rigidly defined.**
b. It should be simple to understand and **easy to calculate.**
c. It should be **based on all the observations** of the data.
d. It should be capable of further **mathematical treatment.**
e. It should be **least affected by the fluctuations** of the sampling.
f. It should not be unduly affected by the extreme values.
g. It should be **easy to interpret.**

MEAN: Mean or average is the most commonly used single descriptive measure of Central Tendency.

Mean is of three types:

a. Arithmetic Mean
b. Geometric Mean
c. Harmonic Mean.

ARITHMETIC MEAN: it simply involves taking the sum of a group of numbers, then dividing that sum by the count of the numbers used in the series.

Q: find arithmetic mean of the following:

X	1	2	3	4	5	6	7	8	9
F	5	10	8	5	3	2	8	6	3

Q: find arithmetic mean of the following:

C.I.	0-20	20-40	40-60	60-80
Freq	5	8	10	7

COMBINED ARITHMETIC MEAN

$X = n_1 X_1 + n_2 X_2 / n_1 + n_2$

Q: The average marks of a group of 50 students in Mathematics are 60 and for other group of 50 students, the average marks are 90. Find the average marks combined group of 100 students.

Mail: banishagupta2099@gmail.com
Website: https://iibf.info

$X = n_1 X_1 + n_2 X_2 / (n_1 + n_2) = 50 \times 60 + 50 \times 90 / 100 = 75$

MERITS OF ARITHMETIC MEAN

a. It is rigidly defined
b. It is **easy to calculate** and simple to follow
c. It is **based on all the observations**
d. It is determined for almost every kind of data
e. It is finite and indefinite
f. It is readily put to algebraic treatment
g. It is least affected by fluctuations of sampling.

DEMERITS OF ARITHMETIC MEAN

1. It is highly affected by extreme values.
2. It **cannot average the ratios** and percentages properly.
3. It is not an appropriate average for highly skewed distribution.
4. It cannot be computed accurately **if any item is missing.**
5. The mean sometimes does not coincide with any of the observed value.
6. Mean cannot be calculated when open-end class intervals are present in the data

GEOMETRIC MEAN: GM is the average value/ mean which measures the central tendency of the set of numbers by taking the root of the product of their values.

- Geometric mean takes into account the compounding effect of the data that occurs from period to period.
- Geometric mean is always less than Arithmetic Mean and is calculated only for positive values.

$$GM\ of\ ungrouped\ data = n\sqrt{(x_1.x_2.x_3 \ldots\ldots.x_n)}$$

$$GM\ of\ grouped\ data = n\sqrt{(x_1^{f1}.x_2^{f2}.x_3^{f_3} \ldots\ldots.x_n^{f_n})}$$

Q: find geometric mean of the following:

X	**1**	**2**	**3**	**4**
F	**5**	**3**	**2**	**4**

Applications

1. It is used in stock indexes.
2. It is used to calculate the annual return on the portfolio.
3. It is used in finance to find the average growth rates which are also referred to the compounded annual growth rate.
4. It is also used in studies like cell division and bacterial growth, etc.

Merits of Geometric Mean

1. It is useful in the **construction of index numbers**.
2. It is **not much affected by the fluctuations** of sampling.
3. It is **based on all the observations.**

Demerits of Geometric Mean

1. It **cannot be easily understood**.
2. It is **relatively difficult to compute** as it requires some special knowledge of logarithms.
3. **It cannot be calculated when any item or value is zero or negative**.

HARMONIC MEAN: It is the reciprocal of the arithmetic mean of reciprocals of the observations.

Arithmetic mean is appropriate measure of central tendency when the values have the same units whereas the Harmonic mean is appropriate measure of central tendency when the values are the ratios of two variables and have different measures.

Harmonic mean is used to calculate the average of ratios or rates.

$$H.M. ungrouped\ data = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_n}}$$

Q: X = 1, 2, 3, 4, 5. Find out harmonic mean?

Q: find harmonic mean of the following:

X	**1**	**2**	**3**	**4**
F	**5**	**3**	**2**	**4**

$$H.M. grouped\ data = \frac{n}{\frac{f_1}{x_1} + \frac{f_2}{x_2} + \cdots + \frac{f_n}{x_n}}$$

APPLICATIONS

- It is used in finance to find average of different rates.
- **It can be used to calculate quantities such as speed**. This is because speed is expressed as a ratio of two measuring units such as km/hr.

Mail: banishagupta2099@gmail.com
Website: https://iibf.info

ARITHMETIC, GEOMETRIC AND HARMONIC MEAN

The arithmetic mean is appropriate if the **values have the same units**, whereas the geometric mean is appropriate if the **values have different units** and **harmonic mean is appropriate if the data values are ratios of two variables with different measures, called rates.**

Arithmetic Mean > Harmonic Mean > Geometric Mean

$$A.M. \times H.M. = (G.M.)^2$$

Q: Find the Harmonic mean of two numbers a and b, if their Arithmetic mean is 16 and Geometric mean is 8.

MEDIAN AND QUARTILES

- The median is the middle value of a distribution
- the number of observations above it is equal to the number of observations below it.
- Observations are arranged either in ascending order or descending order of their magnitude.
- Median is a position average whereas the arithmetic mean is a calculated average.

Median of odd Numbered observations: $= \left(\frac{n+1}{2}\right)^{th} observation$

Median of even Numbered observations:

$$= [\left(\frac{n}{2}\right)^{th} observation + \left(\frac{n+1}{2}\right)^{th} observation)]/2$$

Median of grouped Data

$$Median = I_1 + \frac{(I_2 - I_1)\left(\frac{N}{2} - cf\right)}{f}$$

Q: Calculate median of following data:

C.I.	20-30	30-40	40-50	50-60	60-70
Freq	16	18	20	32	14

CI	Frequency	Class mark	Less than CF
20-30	16	25	16
30-40	18	35	34
40-50	20	45	54
50-60	32	55	86
60-70	14	65	100

$$Median = 40 + \frac{(50 - 40)\left(\frac{100}{2} - 34\right)}{20}$$

Q: Find the median

X	1	2	3	4	5	6	7	8
F	5	10	8	5	3	2	8	6

Mail: banishagupta2099@gmail.com
Website: https://iibf.info

X	F	Less than CF
1	5	5
2	10	15
3	8	23
4	**5**	**28**
5	3	31
6	2	33
7	8	41
8	6	47

N = 47/2 = 23.5

Q: Find the median

C.I.	5-9	10-14	15-19	20-24	25-29
Freq	16	18	20	32	14

C.I	New C.I.	F	Less than CF
5-9	4.5-9.5	16	16
9-14	9.5-14.5	18	34
14-19	14.5-19.5	20	54
19-24	19.5-24.5	32	86
24-29	24.5-29.5	14	100

N/2 = 100/2 = 50, Median Class = 14.5-19.5

$$Median = 14.5 + \frac{(19.5 - 14.5)\left(\frac{100}{2} - 34\right)}{20}$$

QUARTILES

- A quartile represents the **division of data into four equal parts**.
- First, second intervals are based on the data values and third their relationship to the total set of observations.
- By dividing the distribution into four groups, the quartile calculates the range of values above and below the mean.
- A quartile divides data into three points - the lower quartile Q1, the median Q2, and the upper quartile Q3, to create four dataset groupings.

Q: Find the three quartile for the following data:

X	10-15	15-20	20-25	25-30	30-35	35-40	40-45	45-50	50-55
F	14	16	18	12	36	24	30	22	28

CI	F	Less than CF
10-15	14	14
15-20	16	30
20-25	18	48
25-30	12	60
30-35	36	96
35-40	24	120
40-45	30	150
45-50	22	172
50-55	28	200

N/4 = 200/4 = 50 [for Q_1] Median Class → 25-30

$$Q_1 = 25 + \frac{(30-25)\left(\frac{200}{4} - 48\right)}{12} = 25.83$$

N/2 = 200/2 = 100 [for Q_2] Median Class → 35-40

$$Q_2 = 35 + \frac{(40-35)\left(\frac{200}{2} - 96\right)}{24} = 35.83$$

3N/4 = 600/4 = 150 [for Q_3] Median Class → 45-50

$$Q_3 = 45 + \frac{(50-45)\left(\frac{3 \times 200}{4} - 150\right)}{22} = 45$$

MERITS OF MEDIAN

a. It is rigidly defined.
b. It is not affected by extreme values.
c. Even if the extreme values are not known, median can be calculated if the number of items is known.

DEMERITS OF MEDIAN

a. It is not based on all observations.
b. It is affected by sampling fluctuations.
c. It is not capable of further algebraic treatment.

MODE

- The mode of a set of numbers is that number, which occurs more number of times than any other number in the set.
- It is the most frequently occurring value.
- If two or more values occur with equal or nearly equal number of times, then the distribution is said to have two or more modes.

- In case, there are three or more modes and the distribution or data set is said to be **multimodal**.

Q: Find mode for the data: 34, 67, 45, 36, 34, 32, 12, 19, 26, 36, 43, 34.

Mode of grouped Data.

$$Mode = i_1 + \frac{(i_2 - i_1)(f_1 - f_0)}{2f_1 - f_0 - f_2}$$

Q: Find the Mode

C.I.	5-9	10-14	15-19	20-24	25-29
Freq	16	18	20	32	14

Highest frequency is 32 →Modal class is 20-24

$i_2 = 24, i_1 = 20$

$f_0 = 20, f_1 = 32, f_2 = 14$

$$Mode = 20 + \frac{(24 - 20)(32 - 20)}{2 \times 32 - 20 - 14}$$

MERITS OF MODE

a. It is easy to calculate and understand.

b. It is not affected much by sampling fluctuations.

c. It is **not necessary to know all items**. Only the point of maximum concentration is required.

DEMERITS OF MODE

a. It is ill defined as it is not based on all observations.
b. It is **not capable of further algebraic treatment**.
c. It is **not a good representative.**

RELATIONSHIP AMONG MEAN, MEDIAN AND MODE

Mode = 3 Median - 2 Mean

MEASURES OF DISPERSION

- A single value that attempts to describe a set of data by identifying the central position within the set of data is called measure of central tendency.
- Measure of Dispersion is another property of data which establishes the degree of variability or the spread out or scatter of the individual items and their deviation from (or the difference with) the averages or central tendencies.
- A collection of measurements known as dispersion can be used to determine the quality of the data in an objective and quantitative manner.

VARIOUS MEASURES OF DISPERSION ARE GIVEN BELOW

4 ABSOLUTE MEASURES OF DISPERSION

a. Range
b. Quartile Deviation
c. Mean Deviation
d. Standard Deviation

4 RELATIVE MEASURES OF DISPERSION

a. Coefficient of Range
b. Coefficient of Quartile Deviation
c. Coefficient of Mean Deviation
d. Coefficient of Variation

CHARACTERISTICS

a. It should be rigidly defined.
b. It should be based on all observations.
c. It should be easy to calculate and understand.
d. It should be capable of further algebraic treatment.
e. It should not be affected much by sampling fluctuations.

RANGE AND COEFFICIENT OF RANGE

Range: It is the simplest absolute measure of dispersion.

Range (R) = Maximum - Minimum

Coefficient of Range = (Max - Min)/ (Max + Min)

Q: Find range and coefficient of range for: 12, 16, 14, 18, 22.

Range and Coefficient of Range are used to measure the spread in Quality Control, Fluctuations in the Share Prices, in Weather Forecasts:

MERITS OF RANGE

a. It is **easy** to understand.
b. It is **easy to calculate.**

DEMERITS OF RANGE

a. It is not based on all observations.

b. It does not have sampling stability. A single observation may change the value of range.

c. *As the amount of data increases, range becomes less satisfactory.*

QUARTILE DEVIATION AND COEFFICIENT OF QUARTILE DEVIATION

It is the mid-point of the range between two quartiles. Quartile Deviation is defined as

QD = $(Q_3 - Q_1) /2$

Where $Q_1 = 1^{st}$ quartile and $Q_3 = 3^{rd}$ quartile.

Co-efficient of QD = $(Q_3 - Q_1)/ (Q_3 + Q_1)$

MERITS OF QUARTILE DEVIATION

a. It is easy to calculate and understand.

b. It is not affected by extreme values.

DEMERITS OF QUARTILE DEVIATION

a. It is not based on all observations.

b. It is not capable of further algebraic treatment.

c. It is affected by sampling fluctuations.

MEAN DEVIATION AND COEFFICIENT OF MEAN DEVIATION

Mean deviation of a set of observations of a series is the arithmetic mean of all the deviations.

Mail: banishagupta2099@gmail.com
Website: https://iibf.info

It is the deviations from mean when calculated considering their absolute values and are averaged

$$MD = \frac{\sum_{i=1}^{n} |x_i - \bar{x}|}{n}$$

$$coefficient\ of\ mean\ of\ deviation = \frac{MD}{mean}$$

Q: Find Mean Deviation and Coefficient of Mean Deviation from the data: 12, 24, 35, 41, 52, 34, 46, 35.

For Grouped Data:

$$MD = \frac{\sum_{i=1}^{n} f_i |x_i - \bar{x}|}{\sum_{i=1}^{n} f_i}$$

Q: Find the Mean deviation and coefficient of mean deviation:

C.I.	10-20	20-30	30-40	40-50	50-60
Freq	16	18	20	32	14

C.I.	**Freq. (f)**	**Class mark (x)**	$\bar{x}$	**fx**	$f(x - \bar{x})$
10-20	16	15	36	240	336
20-30	18	25	36	450	198
30-40	20	35	36	700	20
40-50	32	45	36	1440	288
50-60	14	55	36	770	266

$$Mean = \frac{\sum fx}{N} = \frac{3600}{100} = 36$$

Mail: banishagupta2099@gmail.com
Website: https://iibf.info

$$MD = \frac{\sum_{i=1}^{n} f_i |x_i - \bar{x}|}{\sum_{i=1}^{n} f_i} = \frac{1108}{100} = 11.08$$

MERITS OF MEAN DEVIATION

a. It is based on all observations.
b. It is easy to understand and also easy to calculate.
c. It is not affected by extreme values.

DEMERITS OF MEAN DEVIATION

a. Mean deviation ignores algebraic signs; hence it is not capable of further algebraic treatment.
b. It is not very accurate measure of dispersion.

STANDARD DEVIATION AND COEFFICIENT OF VARIATION

SD measures the spread or variability of a distribution.

A small standard deviation means a high degree of consistency in the observations as well as homogeneity of the series.

$$SD = \sigma = \sqrt{((x - \bar{x})^2/n)}$$

$$= \sqrt{\frac{x^2}{n} - (\bar{x}^2)}$$

SD of Grouped Data:

$$SD = \sigma = \sqrt{f(x - \bar{x})^2/f)}$$

$$= \sqrt{\frac{fx^2}{N} - \left(\frac{fx}{N}\right)^2}$$

Coefficient of variation = $\frac{\sigma}{\bar{x}}$

Q: Find Standard Deviation and Coefficient of Variation for the following data: 4, 5, 8, 2, 3, 6.

MERITS OF STANDARD DEVIATION

a. It is rigidly defined and has a definite value.
b. It is based on all observations.
c. It is not affected much by sampling fluctuations.

DEMERITS OF STANDARD DEVIATION

a. It is not easy to calculate.
b. It is not easy to understand.
c. It gives more weight to extreme items.

SKEWNESS AND KURTOSIS

- frequency distribution is a said to be symmetrical when the frequency is distributed evenly on either side of an average.
- If frequency distribution is not a symmetrical it is said to skewed.
- Any deviation from symmetry is called skewness.
- skewness is the lack of symmetry.

There are two types of skewness- positive and negative.

If bulk of observations is in the left side of mean and the positive side is longer, it is called positive skewness of the distribution.
In this case, mean and median are greater than mode.

If bulk of observations is in the right side of mean and the negative side is longer, it is called negative skewness of the distribution.
In this case, mean and median are less than mode.

Karl Pearson's measure of skewness is $\beta_1 = \frac{\mu_3^2}{\mu_2^3}$

Where μ_3 = third central moment = $\sum \frac{f(x-\bar{x})^3}{n}$ and μ_2 = second central moment = $\sum \frac{f(x-\bar{x})^2}{n}$

β1 = 0 (symmetrical distribution), β1 > 0 (positive skew), β1 < 0 (negative skew).

KURTOSIS

Kurtosis is all about the tails of the distribution - peakedness or flatness. It is used to describe the extreme values in one versus the other tail. It is the measure of others present in the distribution.

I. The distributions whose peaks are same as of Normal distribution's peak, are called **Mesokurtic.**
II. The distributions whose peaks are higher and sharper than mesokurtic, which means tails are fatter, are called **Leptokurtic** distributions.

III. The distributions whose peaks are lower and shorter than mesokurtic, which means tails are thinner, are called **platukurtic** distributions.

Measure of Kurtosis = β2 = μ_4 / μ^2_2

Where μ_4 = fourth central moment = $\sum \frac{f(x-\bar{x})^4}{n}$

and

μ_2 = second central moment = $\sum \frac{f(x-\bar{x})^2}{n}$

β2 = 0 (Mesokurtic distribution), β2 > 0 (Leptokurtic distribution),

β2 < 0 (Platykurtic distribution),