ambitious baba .com
INDIA'S LARGEST TEST PREP PLATFORM

CAIIB PAPER-1

Module-A Unit-3

Advanced Bank Management (ABM)

CAIIB Paper 1 (ABM) Module A Unit 3: Measures of Central Tendency & Dispersion, Skewness, Kurtosis

Introduction To Measures Of Central Tendency

- Statistical data is first collected (primary or secondary) and then classified into different groups according to common characteristics and presented in a form of a table.
- It is easy for us to study the different characteristics of data from a tabular form.
- Further, graphs and diagrams can also be drawn to convey a better impression to the mind about the data.
- Classified and Tabulated data need to be analysed using different statistical methods and tools and then draw conclusions from it.
- 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.
- According to the statistician, Professor Bowley "Measures of Central Tendency (averages) are statistical constants which enable us to comprehend in single effort the significant of the whole".

The main objectives of Measure of Central Tendency are:

- ✓ To condense data in a single value.
- ✓ To facilitate comparisons between data.

- In other words, the tendency of data to cluster around a central or mid value is called central tendency of data, central tendency is measured by averages.
- There are different types of averages, each has its own advantages and disadvantages.

Requisites of a Good Measure of Central Tendency

- ✓ It should be rigidly defined.
- ✓ It should be simple to understand and easy to calculate.
- ✓ It should be based on all the observations of the data.
- ✓ It should be capable of further mathematical treatment.
- ✓ It should be least affected by the fluctuations of the sampling.
- ✓ It should not be unduly affected by the extreme values.
- ✓ It should be easy to interpret.

Three types of averages are Mean, Median and Mode.

Mean

- Mean or average is the most commonly used single descriptive measure of Central Tendency.
- Mean is simple to compute, easy to understand and interpret.

Mean is of three types:

- ✓ Arithmetic Mean,
- ✓ Geometric Mean
- ✓ Harmonic Mean.

Arithmetic Mean

- The arithmetic mean is the simplest and most widely used measure of a mean, or average.
- 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.

Arithmetic Mean of Ungrouped or Raw Data

$$A = \frac{1}{n}\sum_{i=1}^{n} a_i$$

$$\bar{X} = x1 + x2 + x3 + x4 + ...xn$$

$$\bar{X} = \sum X/n$$

n observations of x.

Example 1: Consider the marks scored by 10 students in Mathematics in a certain examination 35, 30, 18, 15, 40, 30, 52, x, 47, 10. If the arithmetic mean is 30, find the value of x.

$\bar{X}$ = 35 +30 +18 + 15 + 40 +30 +52 + x + 47 + 10/ 10

30 = 277 + x/ 10

300 = 277 + x

X= 23

Arithmetic Mean of Grouped data

- If a variate X take values x1, x2, …, xn with corresponding frequencies f 1, f 2, …, f n respectively, then the arithmetic mean of these values is

$$\bar{X} = \frac{\sum_{i=1}^{n} f_i x_i}{\sum f_i}, x_i$$

X̄ = ∑fx / n

Xi = class marks (mid point) of the class interval for grouped data

Example 2: Find the Arithmetic mean for following:

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

X	F	FX
1	5	5
2	6	12
3	5	15
4	10	40
5	8	40
6	4	24
7	3	21
8	2	16
Total	N= 43	Fx = 173

X̄ = ∑fx / n

= 173/ 43

= 4.02

Combined Arithmetic Mean

If X̄1 and X̄2are the arithmetic mean of two samples of size n1 and n2 respectively then, the Combined arithmetic mean

$$\bar{X} = \frac{n_1 \overline{X_1} + n_2 \overline{X_2}}{n_1 + n_2}$$

Example: The average marks of a group of 100 students in Mathematics are 60 and for other group of 50 students, the average marks are 90. Find the average marks combined group of 150 students.

$$\bar{X} = 100*60 + 50*90/ 100 +50$$
$$= 6000 +4500 / 150$$
$$= 70$$

Example: In private health club, there are 200 members, 100 men, 80 women and 20 children. The average weight of men, women and children are 60 kgs, 50 kgs and 35 kgs respectively. Find the average weight of the combined group.

$n1 = 100, n2 = 80, n3 = 20 \quad x1 = 60, x2 = 50, x3 = 35$

Combined mean =

$\bar{X} = n1\ x1 + n2\ x2 + n3\ x3/\ n1 + n2 + n3$

$= 100*60 + 80*50 + 20*35/ 200$

$= 6000+ 4000 + 700 / 200$

$= 10700/2$

$= 53.5$

Merits of Arithmetic Mean

- It is rigidly defined
- It is easy to calculate and simple to follow
- It is based on all the observations
- It is determined for almost every kind of data
- It is finite and indefinite
- It is readily put to algebraic treatment
- It is least affected by fluctuations of sampling.

Demerits of Arithmetic Mean

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

Geometric Mean

The Geometric Mean (GM) is the average value or 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.

Applications

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

Geometric Mean of Ungrouped or Raw Data

$$\text{G.M.} = \sqrt[n]{x_1 x_2 \ldots x_n} = (x_1 x_2 \ldots x_n)^{\frac{1}{n}} \text{ where } x_1, x_2 \ldots, x_n \text{ are n observations of } x.$$

Example: Find the G.M. of the values 10, 24, 15, and 32.

Given 10, 24, 15, 32

We know that G.M. = 4√ 10*24*15*32

= (10*24*15*32)^1/4

= 115200 ^1/4

= 18.423

Geometric Mean of Grouped or Raw Data

$$\text{G.M.} = \sqrt[n]{x_1^{f_1} x_2^{f_2} \ldots . x_n^{f_n}} = \left(x_1^{f_1} x_2^{f_2} \ldots x_n^{f_n}\right)^{\frac{1}{n}}$$

Example: Find the G.M. for the following data

X	1	2	3	4
F	5	6	5	10

X	F	X^F
1	5	1^5 = 1
2	6	2^6 = 64
3	5	3^5 = 243
4	10	4^10 = 1048576
Total	N = 26	

=26√1*64*243*1048576
= 26√ 1630745394
= 24705

Merits of Geometric Mean

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

Demerits of Geometric Mean

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

Harmonic Mean

- Harmonic Mean is defined as 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. So, generally Harmonic mean is used to calculate the average of ratios or rates.

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.

Harmonic Mean of Ungrouped or Raw data:

$$\text{H.M.} = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}$$ where $x_1, x_2, \ldots\ldots x$ are n observations of x.

H.M = n / (1/x1 + 1/x2 + 1/x3+ ..... 1/y)

Example: Find the HM of the values 10, 24, 15, and 32

= 4/ (1/10+1/24+1/15+1/32)

= 4 / 0.1 + 0.042 + 0.067 + 0.031

= 4/.24

= 16.667

Harmonic Mean of Ungrouped or Raw data:

$$\text{H.M.} = \quad = \frac{n}{\sum_{i=1}^{n} \frac{f_i}{x_i}}$$

N /(F1/X1 + F2/X2 + F3/X3....)

Example: Find the H.M. for the following data

X	1	2	3	4
F	5	6	5	10

X	F
1	5
2	6
3	5
4	10
Total	N = 26

= 26/ (5/1 + 6/2 + 5/3 + 10/4)

= 26/ (5 + 3+ 1.667 + 2.5)

= 26/ 12.167

= 2.137

Comparison between 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. × H.M. = (G.M.)^2**

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

- A.M. = 16 and G.M. = 8
- A.M. × H.M. = G.M^2

- 16 × H.M. = 8^2
- 16 × H.M. = 64
- H.M.= 64/16 = 4

Median And Quartiles

- The median is the middle value of a distribution, i.e., median of a distribution is the value of the variable which divides it into two equal parts.
- It is the value of the variable such that 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 Ungrouped or Raw data

- **The formula to calculate the median of the data is different for odd and even number of observations.**

Median of odd Number of Observations

If the total number of given observations is odd, then the formula to calculate the median for a number of n observations is:

Median = n +1/ 2 th observation

Median of even Number of Observations

If the total number of given observations is even, then the median formula to calculate the median for n number of observations is:

Median = **Median= (n/2)th observation + (n/2+1)th observation / 2**

Example: Find Median of 34, 32, 48, 38, 24, 30, 27, 21, 35.

Arranging the data in ascending order,

21, 24, 27, 30, 32, 34, 35, 38, 48.

n = 9;

Median= (n+1/2) th position

= (9+1/2) the position

= 32

Median of Grouped data:

If variable X takes values X1, X2, X3, X4…..X5 and corresponding frequencies f1, f2, f3, f4,….. Fn respectively, then the median value is given by

$$Median = l_1 + \frac{(l_2 - l_1)\left(\frac{N}{2} - cf\right)}{f}$$

Median class is the class in which the corresponding value of less than cumulative frequency just exceeds the value of N/2.

- l1 = lower limit of the median class,
- l2 = upper limit of the median class
- f = frequency of the median class,
- cf = cumulative frequency of the class preceding the median class,
- N = total frequency.

Example: Find Median for the following data.

Class Interval	20-30	30-40	40-50	50-60	60-70
Frequency	8	26	30	20	16

Class Interval	Frequency	CF
20-30	8	8
30-40	26	34
40-50	30	64
50-60	20	84
60-70	16	100
Total	100	

N/2 = 100/2 = 50
= l1 +[(l2-l1) (N/2 – CF) / f]
= 40 +[10* (50-34) / 30]
= 40 + [10 *16/30]
= 40 + 160/30
= 40 + 5.33
= 45.33

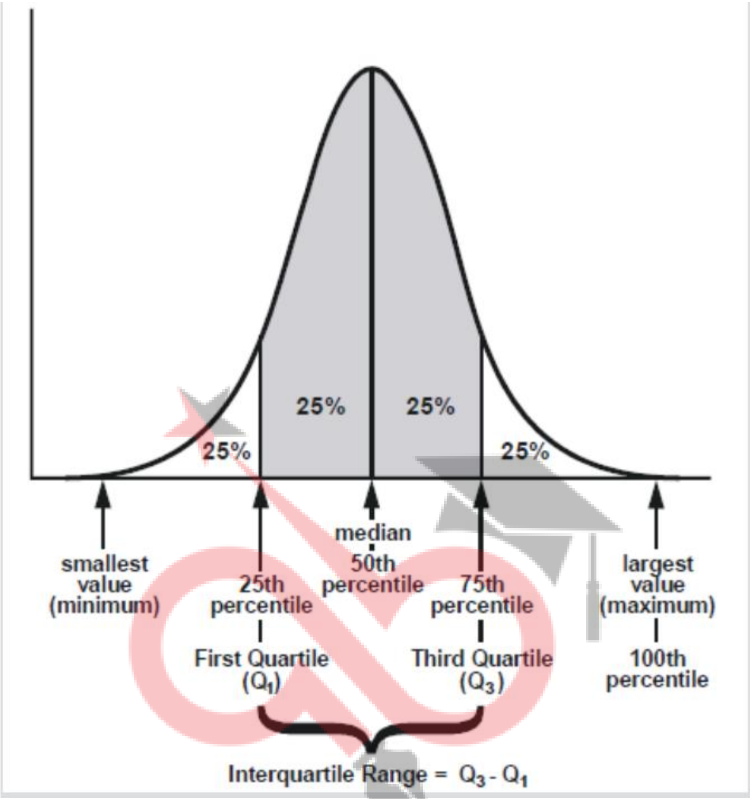
Quartiles

- 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.

The interquartile range is a measure of variability around the median, which is calculated using the quartiles are denoted by Q1, Q2 and Q3



Calculate Q1,Q2 & Q3

Q1 = l1 + (q1-CF)/f (l2-l1) where q1 = N/4

Q2 = l1 + (q2- CF)/f (l2-l1) where q2 = 2N/4

Q3 = l1 + (q3- CF)/f (l2-l1) where q3 = 3N/4

Example: Find the quartiles for the following data

Class Interval	10-15	15-20	20-25	25-30	30-35	35-40	40-45	45-50	50-55
Frequency	12	28	36	50	25	18	16	10	5

Class Interval	Frequency	CF
10-15	12	12
15-20	28	40
20-25	36	76
25-30	50	126
30-35	25	151
35-40	18	169
40-45	16	185
45-50	10	195
50-55	5	200

N/4= 200/4 = 50

Q1 = l1 + (q1-CF)/f (l2-l1)

= 20 + 50-40/36 (5)

= 20 + 10/36 *5

= 21.39

Q2 = l1 + (q2- CF)/f (l2-l1

= 25 + (100-76)/50 (30-25)

= 27.4

AIIB BABA

Q3 = l1 + (q3- CF)/f (l2-l1)

= 30 (150-126) / 25 (35-30)

= 34.8

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.

Mode of Ungrouped or Raw data

Example 22: Find Mode for the data: 23, 25, 20, 23, 26, 21, 27, 28, 30, 27, 23.

Value 23 occurs maximum number of times,

so Mode = 23.

Mode of Grouped data

If a variate X take values x1, x2, x3, x4 with corresponding frequencies f1, f2, f3, f4.... respectively, then the mode is

$$Mode = l_1 + \frac{(l_2 - l_1)(f_1 - f_0)}{2f_1 - f_0 - f_2}$$

Where,

l1 = lower limit of the modal class

l2= per limit of the modal class

f1 = frequency of the modal class

f0 = frequency of the class preceding the modal class

f2 = frequency of the class succeeding the modal class

Example: Find Mode for data

Class Interval	20-30	30-40	40-50	50-60	60-70
Frequency	8	26	30	20	16

Class Interval	Frequency
20-30	8
30-40	26
$\boxed{\text{40-50}}$	$\boxed{30}$
50-60	20
60-70	16
Total	

40 + [10 * 4]/ 60- 26-20
= 40 + 40/14
= 40 + 2.857
= 42.857

$$Mode = l_1 + \frac{(l_2 - l_1)(f_1 - f_0)}{2f_1 - f_0 - f_2}$$

Merits of Mode

- It is easy to calculate and understand.
- It is not affected much by sampling fluctuations.
- It is not necessary to know all items. Only the point of maximum concentration is required.

Demerits of Mode

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

Relationship among Mean, Media and Mode

- **Mode = 3 Median – 2 Mean**

Introduction to 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 a 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.
- The process by which data are scattered, stretched, or spread out among a variety of categories is referred to as dispersion.
- Finding the size of the distribution values that are expected from the collection of data for the particular variable is a part of this process.
- The dispersion of data is a concept in statistics that lets one understand a dataset more simply by classifying individual pieces of data according their own unique dispersion criteria, such as the variance, the standard deviation, and the range.
- 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:

Four Absolute Measures of Dispersion

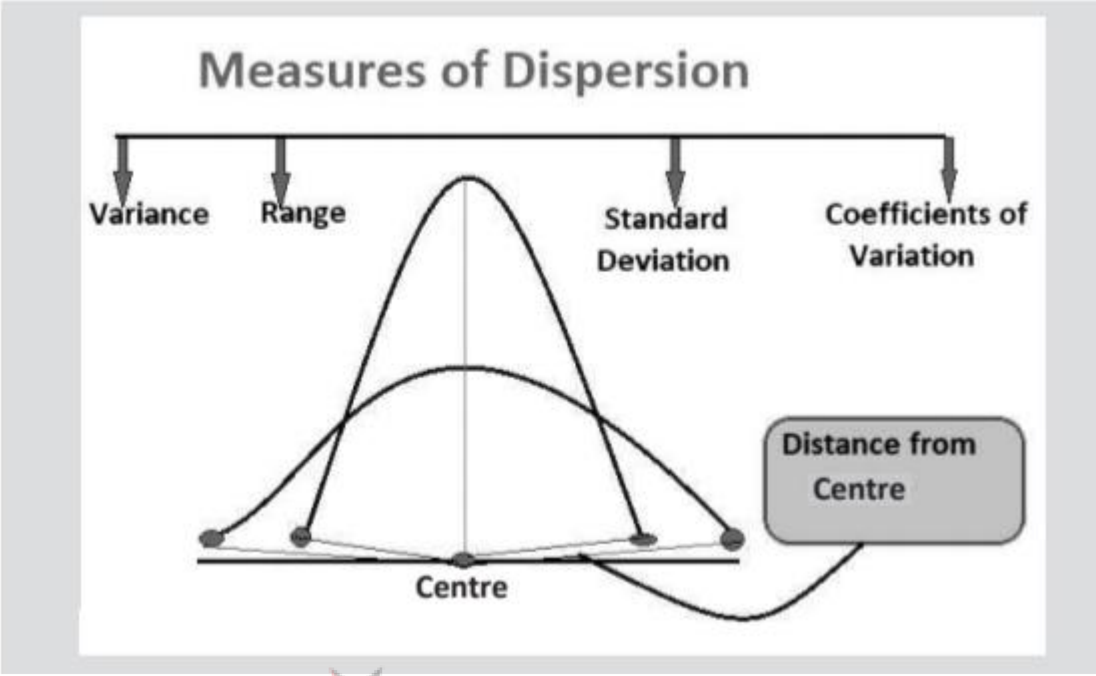
- Range
- Quartile Deviation
- Mean Deviation
- Standard Deviation

Four Relative Measures of Dispersion

- Coefficient of Range
- Coefficient of Quartile Deviation
- Coefficient of Mean Deviation
- Coefficient of Variation

Characteristics of a Good Measure of Dispersion

- It should be rigidly defined.
- It should be based on all observations.
- It should be easy to calculate and understand.
- It should be capable of further algebraic treatment.
- 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)

Example 1 Find the range and coefficient of range of the following items: 18, 15, 20, 17, 22, 16.

- Range = Max – Min = 22 – 15 = 7.
- Coefficient of Range = (Max – Min)/(Max + Min) = 7/37 = 0.19

Note: 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

- It is easy to understand.
- It is easy to calculate.

Demerits of Range

- It is not based on all observations.
- It does not have sampling stability. A single observation may change the value of range.
- 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 = (Q3 – Q1)/2

Where Q1 = 1st quartile and Q 3 = 3rd quartile.

Co-efficient of QD = (Q3 – Q1)/ (Q3 + Q1)

Merits of Quartile Deviation

- It is easy to calculate and understand.
- It is not affected by extreme values.

Demerits of Quartile Deviation

- It is not based on all observations.
- It is not capable of further algebraic treatment.
- 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.
- It is the deviations from mean when calculated considering their absolute values and are averaged.

Mean Deviation (MD) ungrouped data

$$\text{MD (Mean)} = \left(\sum_{i=1}^{n} |x_i - \bar{x}|\right)/n$$

$$\text{Coefficient of Mean Deviation (Mean)} = \frac{\text{MD}(\text{Mean})}{\text{Mean}}$$

$$\mathbf{MD = [(X1 - \bar{X}) + (X2 - \bar{X}) + (X3 - \bar{X}) + \_\_\_\_ + (Xn - \bar{X})] / n}$$

Example: Find Mean Deviation and Coefficient of Mean Deviation

Class Interval	20-30	30-40	40-50	50-60	60-70
Frequency	8	26	30	20	16

Class Interval	Frequency	X	fx	X – X̄ (46)	F (X-X̄)
20-30	8	25	200	21	168
30-40	26	35	910	11	286
40-50	30	45	1350	1	30
50-60	20	55	1100	9	180
60-70	16	65	1040	19	304
Total	100		4600		968

Mean = 4600/100 = 46

- MD (Mean) = 968/100 = 9.68
- Coefficient of Mean Deviation (Mean) = MD (Mean)/ Mean

= 9.68/46 = 0.2104

Merits of Mean Deviation

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

Demerits of Mean Deviation

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

Note: Mean deviation and its coefficient are used in studying economic problems such as distribution of income and wealth in a society.

Standard Deviation And Coefficient Of Variation

- Standard deviation is the most important and commonly used measure of dispersion.
- It 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.

Standard Deviation ungrouped Data

$$\text{SD} = \sigma = \sqrt{\frac{\Sigma\left(x-\bar{x}\right)^2}{n}} = \sqrt{\frac{\Sigma x^2}{n}-\left(\bar{x}\right)^2} \quad \text{where } \bar{x} \text{ is the mean of these observations}$$

Standard Deviation (SD) grouped data

$$\text{SD} = \sigma = \sqrt{\frac{\sum f\left(x-\bar{x}\right)^2}{\sum f}} = \sqrt{\frac{\Sigma fx^2}{N} - \left(\frac{\Sigma fx}{N}\right)^2} \quad \text{where N} = \sum f$$

$$\textbf{Coefficient of Variation = CV} = \frac{\sigma}{\bar{x}} \times 100\%$$

Example: Find Standard Deviation and Coefficient of Variation for the following data: 2, 3, 7, 8, 10.

X	X^2
2	4
3	9
7	49
8	64
10	100
Mean = 30/5=6	226

= √ 226/5 – 6^2
= √ 45.2- 36
= 3.03
Coefficient
= SD/ Mean *100
= 3.03/6 * 100
= 50.5%

$$\text{SD} = \sigma = \sqrt{\frac{\Sigma\left(x-\bar{x}\right)^2}{n}} = \sqrt{\frac{\Sigma x^2}{n} - \left(\bar{x}\right)^2}$$

Example: Find Standard Deviation?

Class Interval	25-30	30-35	35-40	40-45	45-50	50-55
Frequency	30	23	20	14	10	3

Class Interval	Frequency	X	fx	Fx^2
25-30	30	27.5	825	22687.5
30-35	23	32.5	747.5	24293.75
35-40	20	37.5	750	28125
40-45	14	42.5	595	25287.5
45-50	10	47.5	475	22562.5
50-55	3	52.5	157.5	8268.75
Total	N= 100		3550	131225

$$\text{SD} = \sigma = \sqrt{\frac{\sum f\left(x-\bar{x}\right)^2}{\sum f}} = \sqrt{\frac{\Sigma fx^2}{N} - \left(\frac{\Sigma fx}{N}\right)^2}$$

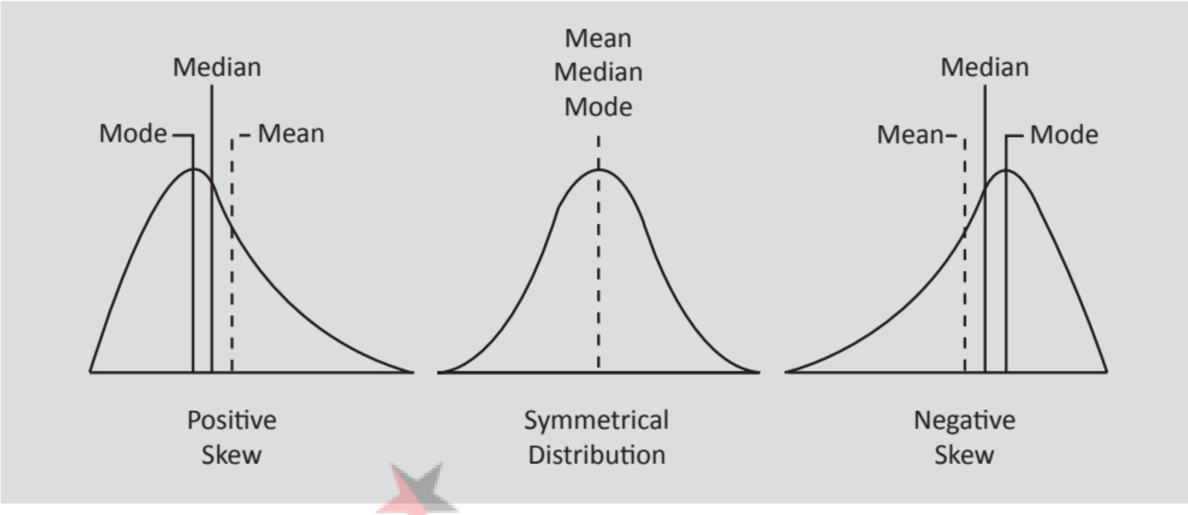
√ 131225/100 – (3550/100)^2
√ 1312.25 – 1260.25
= 7.21

Merits of Standard Deviation

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

Demerits of Standard Deviation

- It is not easy to calculate.
- It is not easy to understand.
- It gives more weight to extreme items.



Skewness And Kurtosis

- Skewness is the degree of distortion from the symmetrical bell curve or the normal distribution.
- It measures the lack of symmetry in data distribution.
- **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.
- mean and median > 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.
- mean and median < mode.

Karl Pearson's measure of skewness is

β1=skewness: = μ3^2/ μ2^3
Where
μ3 = third central moment = ∑f(x-x̄)^ 3/ n
μ2 = second central moment = ∑f(x-x̄)^ 2/ n

- **The direction of skewness is measured by sign of β1, where the sign of β1 is the sign of μ3 .**
- **β1= 0 (symmetrical distribution),**
- **β1> 0 (positive skew),**
- **β1< 0 (negative skew).**

μ1 = First central moment = ∑f(x−x̄)^ 1/ n

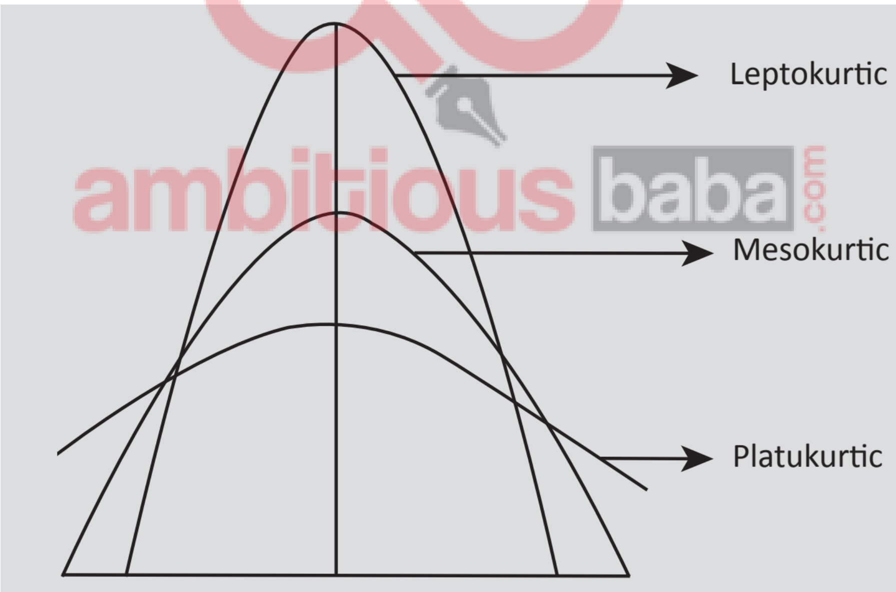
$\mu 2$ = Second central moment = $\sum f(x-\bar{x})^2 / n$

$\mu 3$ = third central moment = $\sum f(x-\bar{x})^3 / n$

$\mu 4$ = Forth central moment = $\sum f(x-\bar{x})^4 / n$

- Kurtosisis 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 actually the measure of outliers present in the distribution.
- The distributions whose peaks are same as of Normal distribution's peak**, are called Mesokurtic.**
- The distributions whose peaks are higher and sharper than mesokurtic, which meanstails are fatter, are called Leptokurtic distributions.
- The distributions whose peaks are lower and shorter than mesokurtic, which means tails are thinner**, are called Platykurtic distributions.**

- **Measure of Kurtosis = $\beta 2 = \mu 4/\mu 2^2$**
- **$\mu 4$ = fourth central moment = $\sum f(x-\bar{x})\ 4/ n$**
- **$\mu 2$ = second central moment = $\sum f(x-\bar{x})\ 2/ n$**
- **$\beta 2 = 0$ (Mesokurtic distribution),**
- **$\beta 2 > 0$ (Leptokurtic distribution),**
- **$\beta 2 < 0$ (Platykurtic distribution).**



- Join CAIIB Telegram Group
- **For Mock test and Video Course Visit**: test.ambitiousbaba.com
- Join Free Classes: **JAIIBCAIIB BABA**

- Download APP For Study Material: Click Here
- Download More PDF

Click here to get Free Study Materials Just by Fill this form