

Chapter 15 – Measures of Central Tendency and Dispersion

FREE Fast Track Lectures (Every day at 10:00 a.m. on YouTube):

<https://www.youtube.com/watch?v=ZZpvRpkgmaE&list=PLAKrxMrPL3fwOSJWxnr8j0C9a4si2Dfdz>

Lecture 1 of Measures of Central Tendency and Dispersion: <https://youtu.be/AEmBanqkVt0>

Lecture 2 of Measures of Central Tendency and Dispersion: <https://youtu.be/C9WkWXODxFs>

Lecture 3 of Measures of Central Tendency and Dispersion: <https://youtu.be/KqanMibPeUs>

Lecture 4 of Measures of Central Tendency and Dispersion: <https://youtu.be/2v5tjtq0DYE>

Unit 1 – Measures of Central Tendency

Categorization of Data

Data can be categorised as follows:

1. Unclassified Data – Individual Series
2. Grouped Frequency Distribution
 - a. Discrete Series
 - b. Continuous Series
 - i. Exclusive Series
 - ii. Inclusive Series

Unclassified Data/ Individual Series

Suppose there are 10 students in a class, and they received marks for their Mathematics Paper held recently. You make a list of the individual marks that every student received:

Student	Marks
Student 1	55
Student 2	65
Student 3	75
Student 4	95
Student 5	80
Student 6	70
Student 7	65
Student 8	60

Student 9	50
Student 10	40

This series is known as the Individual Series. There's only one variable in this series. In the above example, the variable is "Marks".

Individual Series is also known as Unclassified Data.

Grouped Frequency Distribution

Discrete Series

Marks (x_i)	No. of Students (f_i)
50	40
60	10
70	10
80	9
90	11
95	20
Total	100

This type of series is known as a Discrete Series. It comprises of a variable and its frequency. In the above example, "Marks" is the variable, and "No. of Students" is the frequency.

Continuous Series

Marks (Class Interval)	No. of Students (Frequency)
0 – 10	2
11 – 20	5
21 – 30	4
31 – 40	6
41 – 50	10
51 – 60	18
61 – 70	7
71 – 80	15
81 – 90	28
91 – 100	5
Total	100

Continuous Series are of two types:

1. Exclusive Series
2. Inclusive Series

Exclusive Continuous Series

Where the lower limit of a class interval is the same as the upper limit of the preceding class interval, such series is known as Exclusive Series. For example,

Class Interval	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50
Frequency	2	3	4	5	6

In the above table, we can see that the lower limit of the second class interval, (i.e., 10), is the same as the upper limit of the first class interval; the lower limit of the third class interval, (i.e., 20), is the same as the upper limit of the second class interval, and so on.

Inclusive Continuous Series

Where the lower limit of a class interval is NOT the same as the upper limit of the preceding class interval, such series is known as Inclusive Series. For example,

Class Interval	0 – 10	11 – 20	21 – 30	31 – 40	41 – 50
Frequency	2	5	4	6	10

In the above table, we can see that the lower limit of the second class interval, (i.e., 11), is not the same as the upper limit of the first class interval.

Arithmetic Mean

This is the simple average which we've been studying since childhood. If a variable x assumes n

values, $x_1, x_2, x_3, \dots, x_n$, then the arithmetic mean is given by $\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$.

Properties of Arithmetic Mean (AM)

- If all the values assumed by a variable are a constant, say k , then the AM is also k . For example, if the height of every student in a group of 20 students is 180 cm, then the mean height will also be 180 cm (obviously).
- Sum of Deviations from the Mean is always zero, i.e.,
 - For unclassified data: $\sum (x_i - \bar{x}) = 0$
 - For classified data: $\sum f_i (x_i - \bar{x}) = 0$

For example, consider the following unclassified data: 58, 63, 37, 45, 29. Now, $\bar{x} = 46.4$.

Now, the sum of deviations from the mean is calculated as follows:

58	–	46.4	=	11.6
63	–	46.4	=	16.6
37	–	46.4	=	–9.4
45	–	46.4	=	–1.4
29	–	46.4	=	–17.4
Total				<u>0.0</u>

- AM is affected due to a change in origin or scale:
 - Change in Origin:

Suppose you are weighing some students in a class. Suppose the average weight comes to 80 kgs. Afterwards, you find out that the weighing machine was not aligned to zero. It was aligned to 2. This means that when there's no weight on the weighing machine, its pointer, instead of pointing at 0, was pointing at 2. This means that every student's weight is higher by 2 kgs. Now, you'll have to weigh the students again, and calculate the mean again. Instead, you can simply subtract 2 from the original mean 80, and you'll get the corrected mean as 78 kgs.

- Change in Scale:

Suppose you are weighing some students in a class. Suppose the average weight comes to 80 kgs. Afterwards, you realise that the requirement was to calculate the average weight in pounds, and not kgs. Now, you know that 1 kg = 2.205 pounds. Now, you'll again have to weigh all the students in pounds, and then calculate the mean weight. Instead, you may simply multiply 80 with 2.205 and you'll get the corrected mean as 176.4 pounds.

4. Let there be two groups containing n_1 and n_2 observations. Let their arithmetic means be

$$\bar{x}_1 \text{ and } \bar{x}_2 \text{ respectively. Now, the combined AM is given by } \bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2}.$$

Median

For a given set of observations, Median is defined as the middle-most value, when the observations are arranged either in an ascending order or a descending order of magnitude.

Median of Individual Series (Unclassified Data)

Following steps are followed in order to calculate the median of an individual series:

1. Arrange the series in ascending order.
2. Find the rank of the median using the formula $\left(\frac{n+1}{2}\right)$, where n is the total number of observations. In case n is even, this formula will give the mid-point of two terms \rightarrow take the average of those two terms, and that'll be the median. Rank will give us the median term.

Median of Discrete Series

Follow the following steps:

1. Arrange the series in ascending order of the variable.
2. Find cumulative frequency (cf).
3. Find out the rank using the formula $\left(\frac{N+1}{2}\right)$, where N is the total of frequencies.

Median of Continuous Series

Following steps are followed:

1. Make sure that the series is exclusive
2. Arrange the series in ascending order and find cumulative frequency
3. Find the Rank of Median = $\frac{N}{2}$
4. Median (M) = $l + \frac{\text{Rank} - c}{f} \times i$

Where, l = lower limit of Median Class Interval

f = frequency of Median Class Interval

i = Class Interval/Class Size

c = cumulative frequency of the class interval preceding Median Class Interval

Property of Median

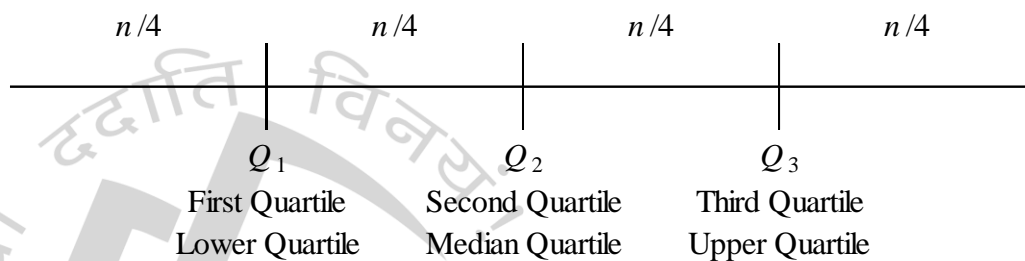
If x and y are two variables related by $y = a + bx$ for any two constants a and b , then the median of y is given by $y_{me} = a + bx_{me}$.

For example, if the relationship between x and y is given by $2x - 5y = 10$, and $x_{me} = 16$, find y_{me} .

$$\text{Now, } 2x - 5y = 10 \Rightarrow 5y = 2x - 10 \Rightarrow y = \frac{2}{5}x - \frac{10}{5} \Rightarrow y = 0.40x - 2 \Rightarrow y = -2 + 0.40x$$

$$\text{Therefore, } y_{me} = a + bx_{me} \Rightarrow y_{me} = -2 + (0.40 \times 16) = -2 + 6.4 = 4.40.$$

Quartile



It is to be noted from the above diagram that the Second Quartile is actually the Median, as it divides the entire series into two equal parts.

Needless to say, that the second quartile is given by the average of the first and the third quartiles. Therefore,

$$\text{Median Quartile} = \frac{\text{First Quartile} + \text{Third Quartile}}{2}$$

Quartile of Individual Series (Unclassified Data)

Following steps are followed to calculate the quartiles of an Individual Series (Unclassified Data):

1. Arrange the series in ascending order.

2. Rank of $Q_1 = \frac{n+1}{4}$

$$\text{Rank of } Q_3 = \frac{3(n+1)}{4}, \text{ where } n \text{ is the total number of observations.}$$

Quartile of Discrete Series

Following steps are followed to calculate the quartiles of a Discrete Series:

1. Arrange the series in ascending order, and find cumulative frequency.

2. Rank of $Q_1 = \frac{N+1}{4}$

$$\text{Rank of } Q_3 = \frac{3(N+1)}{4}, \text{ where } N \text{ is the total of the frequencies.}$$

Quartile of Continuous Series

Following steps are followed to calculate the quartiles of a Continuous Series:

1. Make sure that the series is exclusive.
2. Arrange the series in ascending order and find cumulative frequency.
3. Find the Rank of $Q_1 = \frac{N}{4}$

Rank of $Q_3 = \frac{3N}{4}$, where N is the total of the frequencies.

$$4. Q_1 = l + \frac{\text{Rank} - c}{f} \times i$$

$$Q_3 = l + \frac{\text{Rank} - c}{f} \times i$$

Where, l = lower limit of Class Interval

f = frequency of Class Interval

i = Class Interval/Class Size

c = cumulative frequency of preceding Class Interval

Deciles and Percentiles

Deciles divide a given series into 10 equal parts. Therefore, there are 9 deciles denoted by D_1 to D_9 .

Rank of Decile for an individual series is calculated as follows: $D_1 = \frac{n+1}{10}$, $D_2 = \frac{2(n+1)}{10}$,

$D_3 = \frac{3(n+1)}{10}$, and so on.

Percentiles divide a given series into 100 equal parts. Therefore, there are 99 percentiles denoted by P_1 to P_{99} .

Rank of Percentile for an individual series is calculated as follows: $P_1 = \frac{n+1}{100}$,

$P_2 = \frac{2(n+1)}{100}$, $P_3 = \frac{3(n+1)}{100}$, and so on.

Summary of Partition Values

Partition Value	No. of Parts	No. of Partition Values	Symbol	Rank for Individual Series	Rank for Discrete Series	Rank for Continuous Series
Median	2	1	M	$\frac{n+1}{2}$	$\frac{N+1}{2}$	$\frac{N}{2}$
Quartile	4	3	Q_1 to Q_3	$Q_1 = \frac{n+1}{4}$, $Q_3 = \frac{3(n+1)}{4}$	$Q_1 = \frac{N+1}{4}$, $Q_3 = \frac{3(N+1)}{4}$	$Q_1 = \frac{N}{4}$, $Q_3 = \frac{3N}{4}$

Decile	10	9	D_1 to D_{10}	$D_1 = \frac{n+1}{10}$, $D_5 = \frac{5(n+1)}{10}$ and so on...	$D_1 = \frac{N+1}{10}$, $D_5 = \frac{5(N+1)}{10}$ and so on...	$D_1 = \frac{N}{10}$, $D_5 = \frac{5N}{10}$ and so on...
Percentile	100	99	P_1 to P_{99}	$P_1 = \frac{n+1}{100}$, $P_5 = \frac{5(n+1)}{100}$ and so on...	$P_1 = \frac{N+1}{100}$, $P_5 = \frac{5(N+1)}{100}$ and so on...	$P_1 = \frac{N}{100}$, $P_5 = \frac{5N}{100}$ and so on...

The formula for any partition value of a continuous series is $l + \frac{\text{Rank} - c}{f} \times i$

Mode or Modal Value

Mode is defined as that value which has the highest frequency.

Mode of an Individual Series (Unclassified Data)

Mode of an Individual Series is calculated using the inspection/observation method. We simply have to observe which element occurs the maximum number of times, and that'll give us our mode.

Mode of Discrete Series

In a discrete series, mode is the variable which has the highest frequency.

For example, consider the following series:

X	5	7	9	11
Frequency (f)	25	35	55	98

Since the highest frequency (i.e., 98) is of the variable 11, the mode of the series is 11.

Mode of Continuous Series

Following steps are followed:

1. Locate the class interval with highest frequency. This class interval is known as the Modal Class Interval, and it contains the mode.
2. Calculate Mode using the formula: $\text{Mode} = l + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i$

Here,

l = lower limit of the modal class interval

f_1 = frequency of the modal class interval

f_0 = frequency of the preceding class interval

f_2 = frequency of the succeeding class interval

i = class interval/class size

Properties of Mode

1. Relationship between Mean, Median, and Mode
 - a. For a moderately skewed distribution of data:

Mean – Mode = 3(Mean – Median), or **Mode = 3Median – 2Mean**

b. For a symmetric distribution of data:

Mean = Median = Mode

2. If x and y are two variables related by $y = a + bx$ for any two constants a and b , then the mode of y is given by $y_{mo} = a + bx_{mo}$.

Geometric Mean

Geometric Mean is defined as the n^{th} root of the product of n observations.

Harmonic Mean –

For a given set of non-zero observations, harmonic mean is defined as the reciprocal of the AM of the reciprocals of the observations. Therefore, $HM = \frac{n}{\sum(1/x_i)}$.

Properties of Harmonic Mean

1. If all the observations taken by a variable are constants, say k , then the HM of the observations is also k .
2. The harmonic mean of $1, \frac{1}{2}, \frac{1}{3}, \dots, \frac{1}{n}$ is $\frac{2}{(n+1)}$.
3. If there are two groups with n_1 and n_2 observations and H_1 and H_2 as respective HMs, then the combined HM is given by $\frac{n_1 + n_2}{\frac{n_1}{H_1} + \frac{n_2}{H_2}}$.
4. If rates are given, and an average rate is to be calculated, Harmonic Mean is used.
5. Harmonic Mean is used for calculating Average Speed. Average Speed = $\frac{2xy}{x+y}$.

Relationship between AM, GM, and HM

1. $AM \geq GM \geq HM$
2. $GM^2 = AM \times HM$

Unit 2 – Measures of Dispersion

Range

Individual Series

Range is defined as the difference between the largest and the smallest observation.

Range = Largest Observation – Smallest Observation

Continuous Series

Range = Upper Most Class Boundary – Lower Most Class Boundary

Coefficient of Range

Coefficient of Range = $\frac{L - S}{L + S} \times 100$

Property of Range

Range is not affected with the change in origin, however, when there's a change in the scale, it is affected in the same ratio. In other words, if a and b are two constants related with the variables x and y as $y = a + bx$, then the range of y is given by $R_y = |b| \times R_x$.

Mean Deviation

Mean deviation is defined as the arithmetic mean of the absolute deviations of the observations from an appropriate measure of central tendency.

$$MD_A = \frac{1}{n} \sum |x_i - A|$$

Here, A could either be Mean, or the Median. If A is mean, the above formula gives us the mean deviation about the mean; if A is median, the above formula gives us the mean deviation about the median.

Coefficient of Mean Deviation

Coefficient of Mean Deviation = $\frac{\text{Mean Deviation about } A}{A} \times 100$

Properties of Mean Deviation

1. For a set of observations, the sum of absolute deviations is minimum when the deviations are taken from the median. This property states that $\sum |x_i - \text{Median}|$ is the lowest. In other words, $\sum |x_i - \text{Median}|$ is lower than $\sum |x_i - \text{Any other value}|$.

For example, consider the following set of observations: 2, 4, 6, 8, 20. Clearly, the median is 6. Now, in order to calculate $\sum |x_i - \text{Any other value}|$, let's take the "other value" as 8, and 10.

x	$ x - \text{Median} $, i.e., $ x - 6 $	$ x - 8 $	$ x - 10 $
2	4	6	8
4	2	4	6
6	0	2	4
8	2	0	2
20	14	12	10
Total	$\sum x_i - 6 = 22$	$\sum x_i - 8 = 24$	$\sum x_i - 10 = 30$

From the above table, we can see that the sum of the deviations from the median is minimum.

- Mean Deviation is not affected with the change in origin, however, when there's a change in the scale, it is affected in the same ratio. In other words, if a and b are two constants related with the variables x and y as $y = a + bx$, then the MD of y is given by

$$MD_y = |b| \times MD_x.$$

Standard Deviation

$$SD = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}, \text{ or } SD = \sqrt{\frac{\sum (x_i^2)}{n} - (\bar{x})^2}$$

Sometimes, instead of Standard Deviation, Variance is also used. Variance is nothing but the square of Standard Deviation. Therefore, Variance = $SD^2 = \frac{\sum (x_i - \bar{x})^2}{n}$, or $\frac{\sum (x_i^2)}{n} - (\bar{x})^2$.

Coefficient of Variation

$$\text{Coefficient of Variation (CV)} = \frac{SD}{AM} \times 100$$

Properties of Standard Deviation

- For any two numbers a and b , standard deviation is given by $\frac{|a - b|}{2}$.
- For the first n natural numbers, standard deviation is given by $\sqrt{\frac{n^2 - 1}{12}}$.
- If all the observations assumed by a variable are constant i.e. equal, then the SD is zero. This means that if all the values taken by a variable x is say, k , then $SD = 0$. This result applies to range as well as mean deviation.
- Standard Deviation is not affected with the change in origin.
- When there's a change in the scale, Standard Deviation is affected in the same ratio. In other words, if a and b are two constants related with the variables x and y as $y = a + bx$, then the SD of y is given by $SD_y = |b| \times SD_x$.

6. Consider two groups containing n_1 and n_2 observations, with means \bar{x}_1 and \bar{x}_2 respectively, and standard deviations s_1 and s_2 respectively. The combined SD is given by

$$SD = \sqrt{\frac{n_1 s_1^2 + n_2 s_2^2 + n_1 d_1^2 + n_2 d_2^2}{n_1 + n_2}}$$

where,

$$d_1 = \bar{x}_1 - \bar{x}$$

$$d_2 = \bar{x}_2 - \bar{x}$$

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

Note – When $\bar{x}_1 = \bar{x}_2$, combined SD is given by $\sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2}}$.

7. $SD = \text{Range}/2$

Quartile Deviation

Quartile Deviation is also known as Semi-Inter Quartile Range. It is given by: $QD = \frac{Q_3 - Q_1}{2}$

$$\text{Coefficient of Quartile Deviation} = \frac{Q_3 - Q_1}{Q_3 + Q_1} \times 100$$

Quartile Deviation is not affected with the change in origin, however, when there's a change in the scale, it is affected in the same ratio. In other words, if a and b are two constants related with the variables x and y as $y = a + bx$, then the QD of y is given by $QD_y = |b| \times QD_x$.

Relationship between SD, QD, and MD

- $4SD = 5MD = 6QD$
- $SD = \sqrt{\frac{15}{8}} \times \sqrt{MD \times QD}$ – Question 143
- Ratio between SD, MD, and QD = 15 : 12 : 10