1. Median: The median for a data set is the values that is exactly in the middle position of the list when the data are arranged in order from smallest to largest.

If the number of observations (n) is odd	If the number of observations (n) is even
$Me = \frac{n+1}{2} th\ observation$	$Me = \frac{\left(\left(\frac{n}{2}\right) th\ observation + \left(\frac{n}{2}+1\right) th\ observation\right)}{2}$

Math: Scores of 5 students in an exam have given below: 5, 14, 8, 11, 10. Find median.

<u>**Solution:**</u> First, we rearranged data from smallest to largest, 5, 8, 10, 11, 14

Here, $n = 5$, that is, odd number.

The median can be written as,

$$Me = \frac{(n+1)}{2} th\ observation$$

$$\Rightarrow Me = \frac{5+1}{2} th\ observation = 3rd\ observation = 10$$

Hence, the median mark of all student is 10.

Math: Scores of 10 students in an exam have given below: 27, 14, 92, 5, 68, 31, 83, 10, 45, 77. Find median.

<u>**Solution:**</u> First, we rearranged data from smallest to largest, 5, 10, 14, 27, 31, 45, 68, 77, 83, 92

Here, $n = 10$, that is, even number.

The median can be written as,

$$Me = \frac{\left(\left(\frac{n}{2}\right) th\ observation + \left(\frac{n}{2}+1\right) th\ observation\right)}{2}$$

$$\Rightarrow Me = \frac{5th\ observation + 6th\ observation}{2} = \frac{(31+45)}{2} = 38$$

Hence, the median mark of all student is 38.

Math: Find the median of the below data.

Wealth Status	Poorest	Poorer	Middle	Richer	Richest
Frequency	4	10	17	12	7

Solution: Here,

Wealth Status	**Frequency**	$\boldsymbol{CF_i}$
Poorest	4	4
Poorer	10	14
Middle	17	31
Richer	12	43
Richest	7	50
Total	$n = 50$	

Now, $N/2 = 25$, which lie in wealth index group "Middle". Hence, wealth index group "Middle" is the median.

2. Median for grouped data:

$$Me = L_m + \frac{\frac{N}{2} - F_c}{f_m} \times c$$

Here,

$L_m = Lower\ limit\ of\ median\ class$

$N = Total\ number\ of\ observations/frequency$

$F_c = Cumultative\ frequency\ of\ pre-median\ class$

$f_m = Frequency\ of\ median\ class$

$c = Class\ interval$

Steps of computing median from grouped data:

1. Prepare a less than type cumulative frequency distribution.
2. Determine $\frac{N}{2}$, where N is the total frequency.
3. Locate the median class whose cumulative frequency includes the value of $\frac{N}{2}$.
4. Determine the value of $L_m, F_c, f_m, and\ c$.

<u>Math:</u> Calculate median,

Class	20-30	30-40	40-50	50-60	60-70	70-80	Total
f_i	4	11	9	7	5	4	40

Solution: First, we reconstruct the table,

Class	$\boldsymbol{f_i}$	$\boldsymbol{CF_i}$
20-30	4	4
30-40	11	15
40-50	9	24
50-60	7	31
60-70	5	36
70-80	4	40
Total	40	

Here, $N = 40; \frac{N}{2} = \frac{40}{2} = 20.$

Looking at the cumulative frequency column in the table, we find that the position of 20 which is in the class $(40 - 50)$.

Now, $L_m = 40, N = 40, F_c = 15, f_m = 9, and\ c = 10$

$$\therefore Me = L_m + \frac{\frac{N}{2} - F_c}{f_m} \times c = 40 + \frac{20 - 15}{9} \times 10 = 45.55$$

3. Merits of median:

a) It is rigidly defined.
b) Median is easy to understand and easy to calculate for a non-mathematical person.
c) Since median is a positional average, it is not affected at all by extreme observations.

4. Demerits of median:

a) In case of even number of observations for an ungrouped data, median cannot be determined exactly.
b) Median is relatively less stable than mean.

5. When is the median a more suitable choice than the mean?

The median is more applicable than the mean in situations where the data distribution is skewed or contains outliers. Here are some scenarios where using the median is preferred:

a) ***Skewed Distributions:*** When the data distribution is skewed, meaning that it's not symmetric and has a longer tail on one side, the median can provide a better representation of the "typical" value. The mean can be significantly influenced by the skewed tail, whereas the median is less affected.

b) ***Outliers:*** Outliers are extreme values that can disproportionately impact the mean. In such cases, the median is a better measure of central tendency because it is resistant to the influence of outliers. It reflects the value that falls exactly in the middle of the ordered data set.

c) ***Ordinal Data:*** When working with ordinal data (data with ordered categories but no consistent intervals), the median is often more meaningful than the mean. For example, in survey responses with Likert scales, the median indicates the middle point of respondents' opinions.

6. Mode: The mode of a data set is its most frequently occurring value. Data can have more than one mode. If it has two modes, it is referred to as bimodal, three modes, tri-modal, and the like.

*** Mode is the only one measure of central tendency, which can be computed at all levels of measurement.

*** 3,4,0,2,2,1,2,5,2, 2; the modal value is 2 and the data set is unimodal data.

*** 2,2,2,0,3,4,4,4,5,6; the modal value is 2 and 4, and the data set is bimodal data.

*** 2,3,4,5,1,9,6,10,7; there is no modal value.

*** 1,1,2,2,3,3,4,4; the modal value is 1, 2, 3, and 4, and the data set is multimodal data.

7. Merits of mode:

a) Mode is easy to calculate and understand.
b) Mode is not at all affected by extreme observations and as such is preferred to arithmetic mean while dealing with extreme observations.

8. Demerits of mode:

a) Mode is not rigidly defined.
b) It is not based on all the observations of the series.
c) Mode is not suitable for further mathematical treatment.

Math: The data below represent the sampled frequency distribution of “Programming Languages Known” in CSE department. Find the mode.

Number of Programming Languages Known	**Frequency**
0	5
1	12
2	15
3	18
4+	8
Total	58

Solution: The table reveals that the occurrence of students who know three programming languages is the highest, with a frequency of 18 times. As a result, the modal value for “programming languages known” is determined to be 3.

Math: The data below represent the sampled percentage distribution of “Programming Languages Known” in CSE department. Find the mode.

Number of Programming Languages Known	**Frequency**	**Percentage**
0	5	8.6%
1	12	20.7%
2	15	25.9%
3	18	31.0%
4+	8	13.8%
Total	58	100%

Solution: The table reveals that the occurrence of students who know three programming languages is the highest, with 31%. As a result, the modal value for "programming languages known" is determined to be 3.

9. Mode for grouped data:

$$\boldsymbol{Mode = 3Median - 2Mean}$$

$$Mo = L_o + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times c$$

Here,

$L_o = Lower\ limit\ of\ modal\ class$

$\Delta_1 = Difference\ between\ frequency\ of\ the\ modal\ class\ and\ pre-modal\ class.$

$\Delta_2 = Difference\ between\ frequency\ of\ the\ modal\ class\ and\ post-modal\ class.$

$c = Class\ interval$

Math:

Class	$\boldsymbol{f_i}$
10-20	5
20-30	8
30-40	12
40-50	7
50-60	9

Here, the class with highest frequency is $(30 - 40)$. This is our modal class.

Now, $L_o = 30, \Delta_1 = 12 - 8 = 4, \Delta_2 = 12 - 7 = 5, c = 10$

$$\therefore Mode, Mo = L_o + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times c = 30 + \frac{4}{4+5} \times 10 = 34.44$$

10. Comparison of mean, median and mode: (HOME WORK)

11. Does all data have a median, mode and mean?

Every set of continuous data possesses a median, mode, and mean. Nevertheless, when considering ordinal data, it encompasses solely a median and mode, whereas nominal data solely involves a mode.

	Mean	***Median***	***Mode***
Nominal	No	No	Yes
Ordinal	No	Yes	Yes
Interval	Yes	Yes	Yes
Ratio	Yes	Yes	Yes

12. What is the best measure of central tendency?

There can often be a "best" measure of central tendency with regards to the data you are analyzing, but there is no one "best" measure of central tendency. This is because whether we use the median, mean or mode will depend on the type of data we have (see our types of variables guide), such as nominal or continuous data; whether your data has outliers; and what you are trying to show from your data.

13. If there is outlier in dataset, what is the best indicator of central tendency?

It is usually inappropriate to use the mean in such situations. We would normally choose the median or mode, with the median usually preferred.

14. When is the mean the best measure of central tendency?

The mean is usually the best measure of central tendency to use when our data distribution is continuous and symmetrical (no outlier) and quantitative variable, However, it all depends on what you are trying to show from your data.

15. When is the mode the best measure of central tendency?

The mode is the least used of the measures of central tendency and can only be used when dealing with nominal data. For this reason, the mode will be the best measure of central tendency (as it is the only one appropriate to use) when dealing with nominal data.

16. When is the median the best measure of central tendency?

The median is usually preferred to other measures of central tendency when your data set is skewed (i.e., has outlier observations) or we are dealing with ordinal data. However, the mode can also be appropriate in these situations, but is not as commonly used as the median.

17. What is the most appropriate measure of central tendency when the data has outliers?

The median is usually preferred in these situations because the value of the mean can be distorted by the outliers. However, it will depend on how influential the outliers are. If they do not significantly distort the mean, using the mean as the measure of central tendency will usually be preferred.

Extra:

- Table below shows the frequency distribution of marks in "Statistics" of 50 students of a university.

Scores	0-5	5-10	10-15	15-20	20-25	25-30	30-35
Freq.	3	6	9	9	14	8	1

a) Compute mean, median, and mode.

- Below is the frequency distribution of motor cycles sold at a shopping mall las month.

Sales ('000 BDT)	100-130	130-160	160-190	190-220	220-250	250-280	280-310
Number of m.cycles	10	22	40	70	45	10	3

a) Compute mean, median, and mode.

1. Measures of Location:

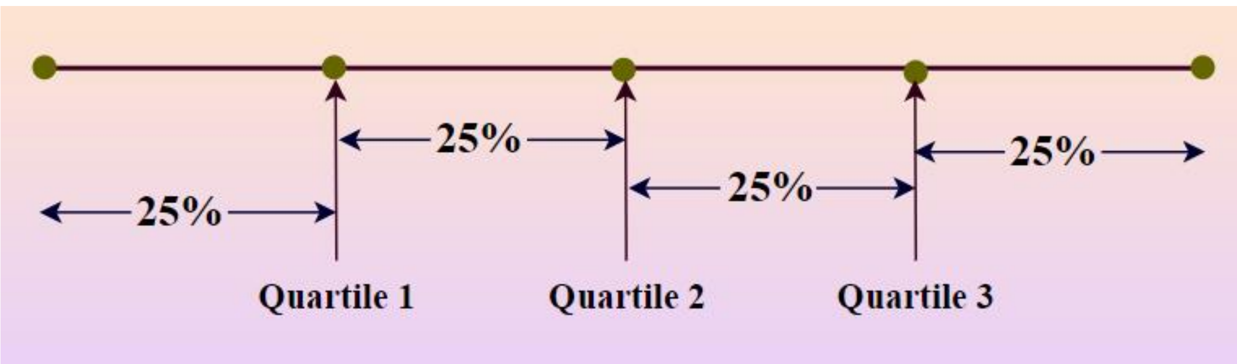
a) Quartiles

b) Deciles

c) Percentiles

2. Quartiles: The quartiles which we denote Q_1, Q_2, and Q_3 divide a data set into four equal parts when the data arranged in order from smallest to largest.

- The first quartile (Q_1) is the value that separates the bottom 25% of the data from the top 75%
- The second quartile (Q_2) is the middle value/ median value and it separates the bottom 50% of the data from the top 50%
- The third quartile (Q_3) is the value that separates the bottom 75% of the data from the top 25%



Steps of getting Quartiles:

a) Arrange data set from smallest to largest

b) Identify the position of $Q_i; i = 1,2,3$ by utilizing the formula,

$$J = \frac{(i \times n)}{4}$$

c) If J is the integer value, then

$$\frac{(J^{th} Observation + (J+1)^{th} Observation)}{2}$$

If J is not integer value, then take the next integer value as position.

Math: Scores from 11 CSE students. Compute the quartiles of the data set and interpret the results,

$$20, 46, 27, 38, 50, 33, 36, 58, 23, 22, 60$$

Solution:

First, we arrange the data set in order from smallest to largest:

$$20, 22, 23, 27, 33, 36, 38, 46, 50, 58, 60$$

Now,

$$\therefore Position\ of\ Q_1 = \frac{(i \times n)}{4}$$

$$\Rightarrow Position\ of\ Q_1 = \frac{(1 \times 11)}{4} = 2.75$$

Since the position value is not integer, then we take the next integer value as position.

$$\therefore Q_1 = 3^{rd} Obervation = 23$$

Interpretation: It's evident that 25% of the students achieve scores of 23 or below, while 75% of the students attain scores above 23.

$$\therefore Position\ of\ Q_2 = \frac{(i \times n)}{4}$$

$$\Rightarrow Position\ of\ Q_2 = \frac{(2 \times 11)}{4} = 5.5$$

Since the position value is not integer, then we take the next integer value as position.

$$\therefore Q_2 = 6^{th} Observation = 36$$

Interpretation: It's evident that 50% of the students achieve scores of 36 or below, while 50% of the students attain scores above 36.

$$\therefore Position\ of\ Q_3 = \frac{(i \times n)}{4}$$

$$\Rightarrow Position\ of\ Q_3 = \frac{(3 \times 11)}{4} = 8.5$$

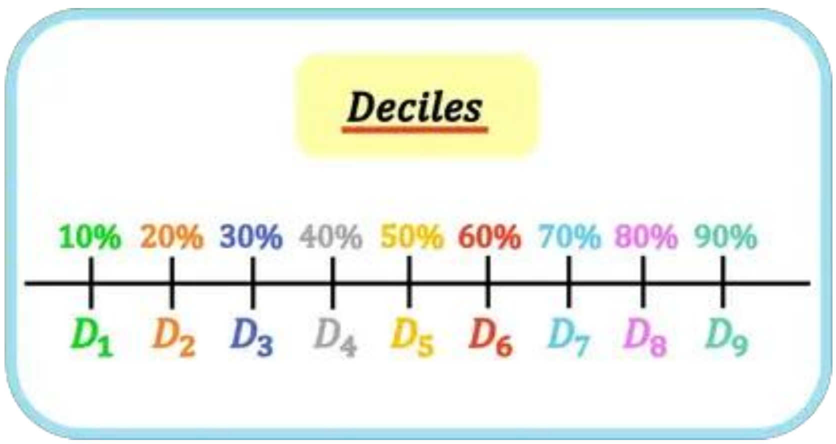
Since the position value is not integer, then we take the next integer value as position.

$$\therefore Q_3 = 9^{th} Observation = 50$$

Interpretation: It's evident that 75% of the students achieve scores of 50 or below, while 25% of the students attain scores above 50.

3. Deciles: The deciles which we denote $D_1, D_2, \dots, D_9$ divide a data set into ten equal parts when the data arranged in order from smallest to largest.

- The first decile (D_1) is the value that separates the bottom 10% of the data from the top 90%.
- The second decile (D_2) is the value that separates the bottom 20% of the data from the top 80%.
- Proceeding in this way, the ninth decile (D_9) separates the bottom 90% value from the top 10%.



Steps of getting Deciles:

a) Arrange data set from smallest to largest

b) Identify the position of $D_i; i = 1,2,3, \dots, 9$ by utilizing the formula,

$$J = \frac{(i \times n)}{10}$$

c) If J is the integer value, then

$$\frac{(J^{th} Observation + (J+1)^{th} Observation)}{2}$$

If J is not integer value, then take the next integer value as position.

Math: Scores from 11 CSE students. Compute the value of D_1, D_5, D_9 of the data set and interpret the results,

$$20, 46, 27, 38, 50, 33, 36, 58, 23, 22, 60$$

Solution:

First, we arrange the data set in order from smallest to largest:

$$20, 22, 23, 27, 33, 36, 38, 46, 50, 58, 60$$

Now,

$$\therefore Position\ of\ D_1 = \frac{(i \times n)}{10}$$

$$\Rightarrow Position\ of\ D_1 = \frac{(1 \times 11)}{10} = 1.1$$

Since the position value is not integer, then we take the next integer value as position.

$$\therefore D_1 = 2^{nd} Obervation = 22$$

Interpretation: It's evident that 10% of the students achieve scores of 22 or below, while 90% of the students attain scores above 22.

$$\therefore Position\ of\ D_5 = \frac{(i \times n)}{10}$$

$$\Rightarrow Position\ of\ D_5 = \frac{(5 \times 11)}{10} = 5.5$$

Since the position value is not integer, then we take the next integer value as position.

$$\therefore D_5 = 6^{th} Observation = 36$$

Interpretation: It's evident that 50% of the students achieve scores of 36 or below, while 50% of the students attain scores above 36.

$$\therefore Position\ of\ D_9 = \frac{(i \times n)}{10}$$

$$\Rightarrow Position\ of\ D_9 = \frac{(9 \times 11)}{10} = 9.9$$

Since the position value is not integer, then we take the next integer value as position.

$$\therefore D_9 = 10^{th} Observation = 58$$

Interpretation: It's evident that 90% of the students achieve scores of 58 or below, while 10% of the students attain scores above 58.

4. Percentiles: The percentiles which we denote $P_1, P_2, \dots, P_{99}$ divide a data set into hundred equal parts when the data arranged in order from smallest to largest.

- The first percentile (P_1) is the value that separates the bottom 1% of the data from the top 99%.
- Proceeding in this way, the ninety ninth percentile (P_{99}) separates the bottom 99% of the data from the top 1%.

Steps of getting Percentiles:

a) Arrange data set from smallest to largest

b) Identify the position of $P_i; i = 1,2,3, \dots, 99$ by utilizing the formula,

$$J = \frac{(i \times n)}{100}$$

c) If J is the integer value, then

$$\frac{(J^{th} Observation + (J+1)^{th} Observation)}{2}$$

If J is not integer value, then take the next integer value as position.

Math: Scores from 11 CSE students. Compute the value of P_{25} of the data set and interpret the results,

$$20, 46, 27, 38, 50, 33, 36, 58, 23, 22, 60$$

Solution:

First, we arrange the data set in order from smallest to largest:

$$20, 22, 23, 27, 33, 36, 38, 46, 50, 58, 60$$

Now,

$$\therefore Position\ of\ P_{25} = \frac{(i \times n)}{100}$$

$$\Rightarrow Position\ of\ P_{25} = \frac{(25 \times 11)}{100} = 2.75$$

Since the position value is not integer, then we take the next integer value as position.

$$\therefore P_{25} = 3^{rd} Obervation = 23$$

Interpretation: It's evident that 25% of the students achieve scores of 23 or below, while 75% of the students attain scores above 23.

$$Q_i = L_Q + \frac{\frac{i \times N}{4} - F_{Qc}}{f_Q} \times c$$

$$D_i = L_D + \frac{\frac{i \times N}{10} - F_{Dc}}{f_D} \times c$$

$$P_i = L_P + \frac{\frac{i \times N}{100} - F_{Pc}}{f_P} \times c$$

*** A social organization publishes information on the TV-viewing habits of Bangladeshi garments workers in a report on Television. A sample of 100 garments workers yielded the monthly viewing times, in hours, displayed on the table,

TV-viewing times (in hours)	f_i
<50	8
50-60	14
60-70	35
70-80	20
80-90	15
≥90	8

a) Determine $Q_1, Q_2, and\ Q_3$ with interpretation. [Ans: 60.9, 68, and 79]

b) Determine $D_1, D_5, and\ D_3$ with interpretation. [Ans: 51.43, 68, and 88.67]

c) Determine $P_{30}, and\ P_{30}$ with interpretation. [Ans: 62.29, and 68]

Sample answer:

TV- viewing times (in hours)	Frequency (Number of workers)	Cumulative Frequency
40- 50	8	8
50 - 60	14	22
60 - 70	35	57
70 - 80	20	77
80 - 90	15	92
90 and more.	8	100

Determine quartiles and interpret the results.

Solution Number of observations, $N = 100$

Computation of Q_1:

$$\frac{1 \times N}{4} = \frac{1 \times 100}{4} = 25$$

We see that the first quartile class is 60-70.

$$Q_1 = L_1 + \frac{\frac{1 \times N}{4} - F_c}{f_Q} \times c$$

$$= 60 + \frac{25 - 22}{35} \times 10 = 60 + \frac{3}{35} \times 10 = 60.9$$

Where, $L_1 = 60$

$F_c = 22$

$f_Q = 35$

$c = 10$