Day 3
Data Management and Analysis

Session 1
Measures of Central Tendency

Session Outcome

After completing this session, researchers will be able to

- Understand the Concept of Measures of Central Tendency and their Importance
- Understand Mean, Median, and Mode as Measures of Central Tendency.
- Explain the Formulas and Procedures to Calculate the Mean, Median, and Mode.

Session Outline

- Definition and Importance of Measures of Central Tendency
- Mean, Median, and Mode (for Grouped and Ungrouped data)
- Advantages and Disadvantages of Different Measures

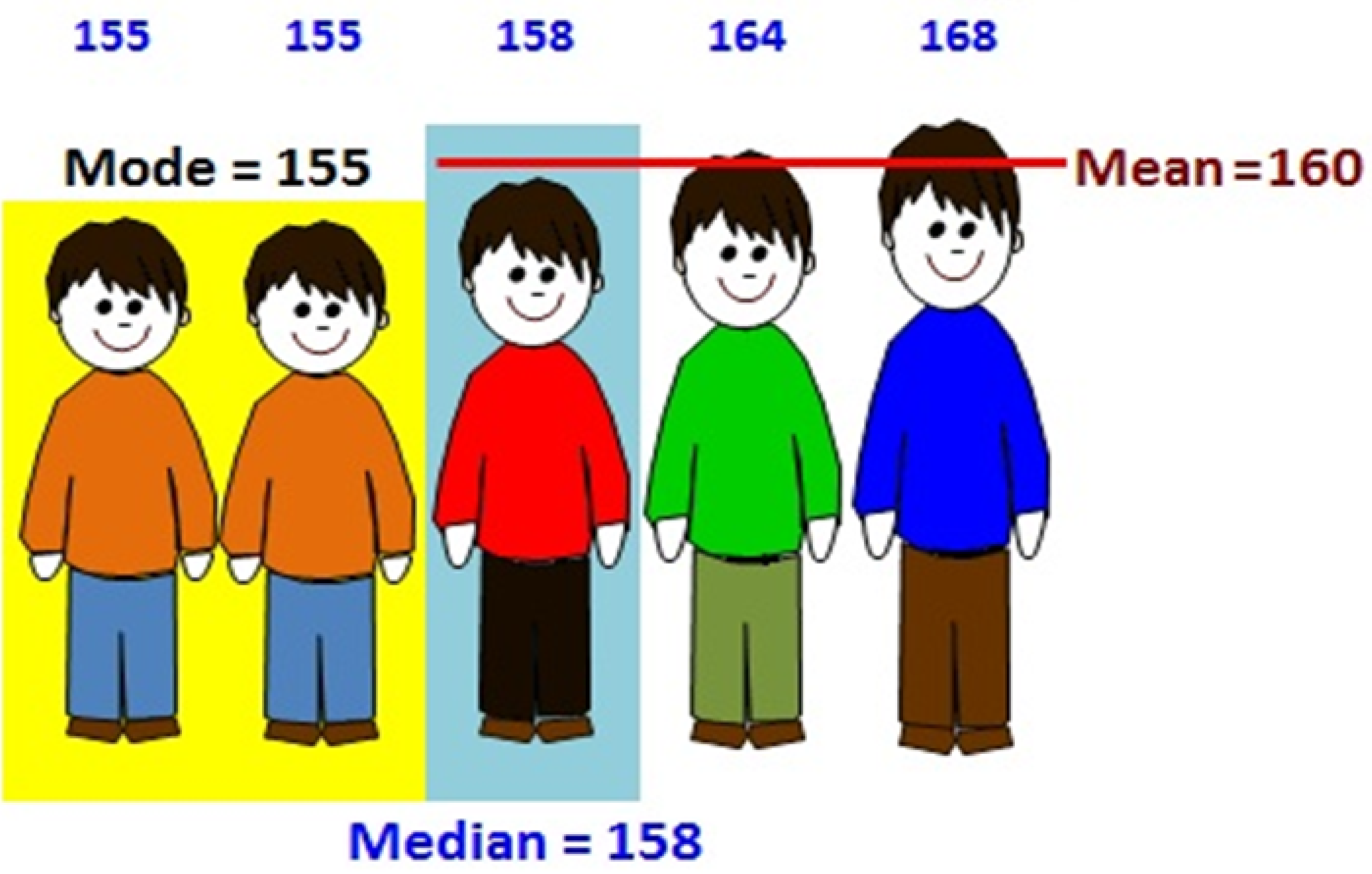
Measures of Central Tendency

Measures of central tendency are statistical measures, that represent a single value that attempts to describe a set of data by identifying the central position within that set of data.

Example:

A primary education researcher studies the physical development of boys in a school by measuring their heights. They calculate the average height, which provides valuable information about the typical height of students in that grade.

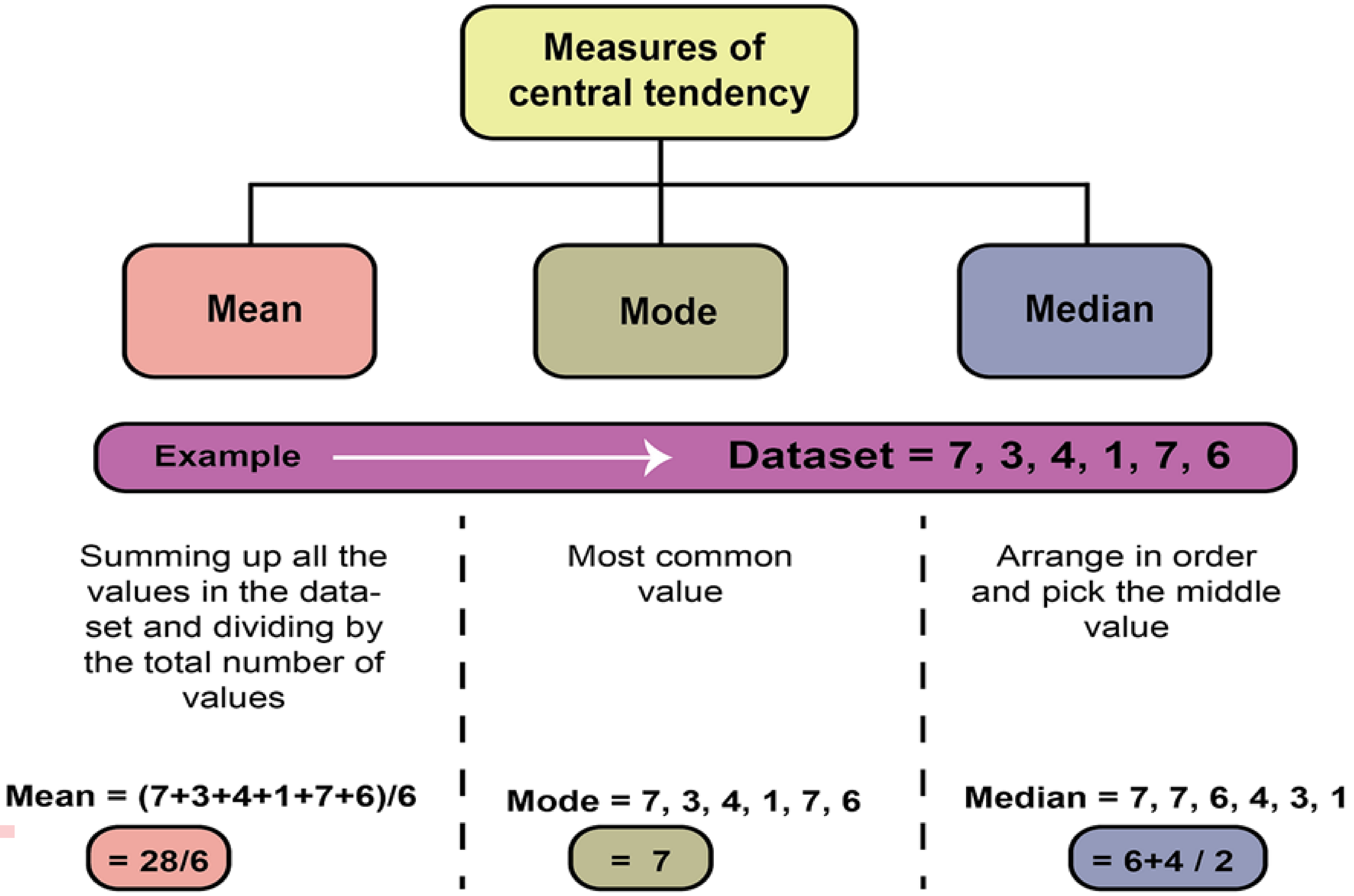
Measures of Central Tendency



Importance of measures of central tendency:

- Student Performance Evaluation
- Identifying Areas of Strength and Weakness
- Setting Benchmarks and Goals
- Monitoring Progress & Resource Allocation
- Data-Driven Decision-Making
- Communicating with Stakeholders

Types of Measures of Central Tendency



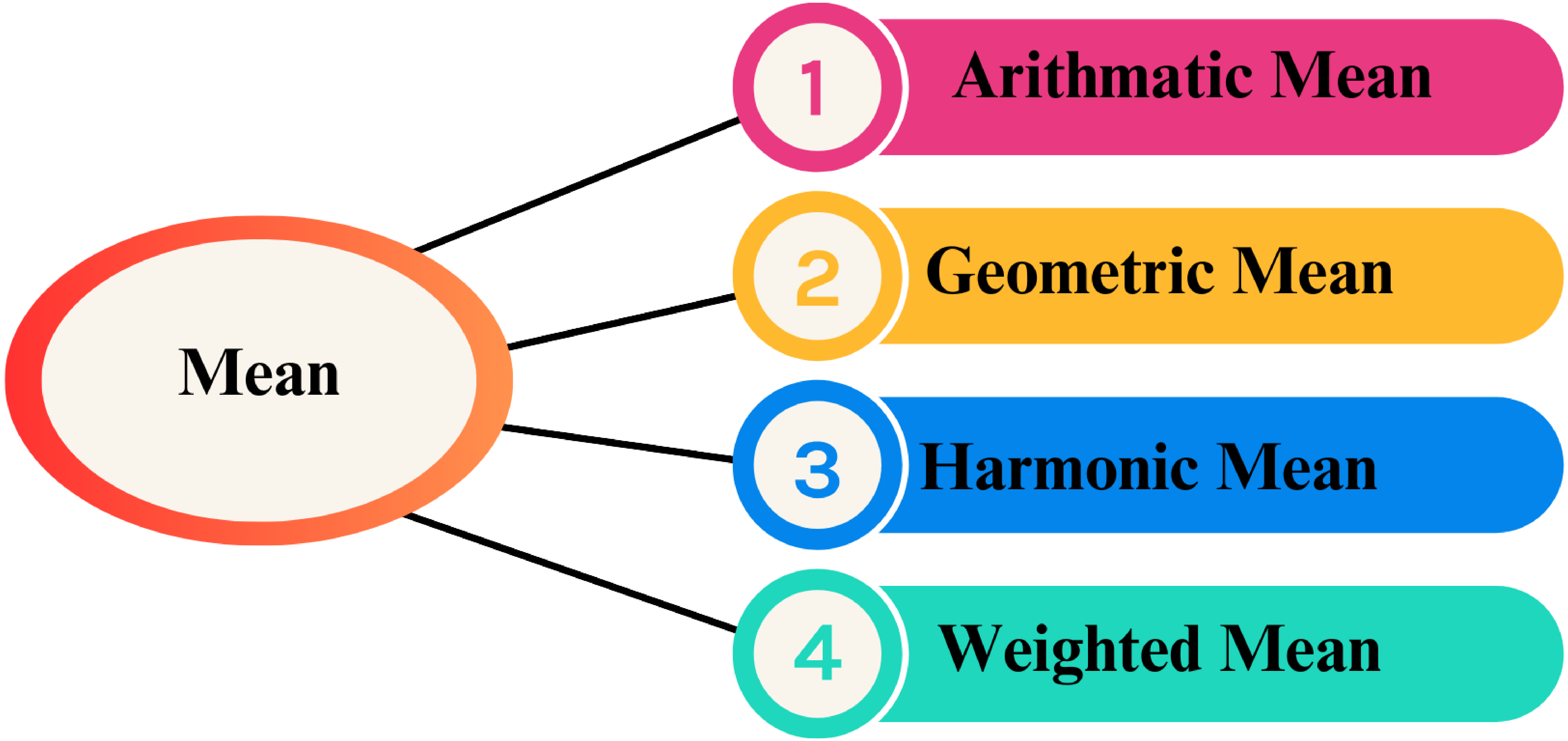
Types of Measures of Central Tendency

Here are the main types of measures of central tendency:

- **Mean**: The mean in mathematics and statistics is the average value of a set of numbers.
- **Median**: The median is the middle value in a data set when the values are arranged in ascending or descending order.
- **Mode**: The mode is the value that appears most frequently in a data set.

Mean as a Measure of Central tendency

The mean is one of the most common measures of central tendency used in statistics. There are four types of mean and they are,



Mean as a Measure of Central tendency

Arithmetic Mean (AM): The arithmetic mean is a central tendency measure that represents the average of a set of numerical values, calculated by summing all values in the dataset.

1. Arithmetic Mean (AM) for Ungrouped Data:

For ungrouped data, each value is treated individually, and the mean is calculated straightforwardly using the formula:

Mathematically, the arithmetic mean $\bar{x}$ of a dataset $\mathbf{x_1}, \mathbf{x_2}, \ldots \mathbf{x_n}$ is calculated as:

$$AM(\overline{x}) = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Where:

- $\mathbf{x_1}, \mathbf{x_2}, \ldots \mathbf{x_n}$ are the individual values in the dataset
- n is the total number of values in the dataset.

Mean as a Measure of Central tendency

2. Arithmetic Mean (AM) for Grouped Data:

For grouped data, where values are organized into intervals or classes, the calculation involves using the midpoint of each class interval. The formula is:

$$AM(\overline{x}) = \frac{\sum_{i=1}^{k} f_i m_i}{N}$$

Where:

- f_i represents the frequency of the i^{th} class interval.
- m_i represents the midpoint of the i^{th} class interval.
- N is the total frequency, which is the sum of all frequencies.
- k is the total number of class intervals.

Mean as a Measure of Central tendency

Example: Arithmetic Mean of Test Scores

Suppose we have a class of 25 primary school students, and we want to calculate the arithmetic mean of their test scores. Here are the test scores:

80	85	75	90	92	88	82	78	85	88	72	95	85	90
84	80	82	88	86	92	90	75	80	85	88			

Arithmetic Mean for Ungrouped Data:

Using the formula for ungrouped data, we sum up all the individual test scores and divide by the total number of students:

$$AM(\overline{x}) = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$= \frac{2073}{25} = 82.92$$

So, the arithmetic mean test score for the class is approximately 82.92.

Mean as a Measure of Central tendency

Arithmetic Mean for Grouped Data: Now, let's say we want to group the test scores into intervals (e.g., 70-79, 80-89, 90-100) and calculate the arithmetic mean.

Suppose the frequency distribution of the test scores is as follows

Test Score Range	Frequency	Mid-point
70-79	3	74.5
80-89	15	84.5
90-100	7	94.5

Mean as a Measure of Central tendency

To calculate the arithmetic mean for grouped data, we'll find the midpoint of each interval and then use the formula:

$$Arithmetic\ Mean = \frac{\sum_{i=1}^{k} f_i m_i}{N}$$

$$Arithmetic\ Mean = \frac{(3 \times 74.5) + (15 \times 84.5) + (7 \times 94.5)}{25}$$

$$Arithmetic\ Mean = \frac{2152.5}{25}$$

$$Arithmetic\ Mean = 86.1$$

Therefore, the arithmetic mean test score for the class, when grouped into intervals, is approximately 86.1.

Mean as a Measure of Central tendency

Geometric Mean (GM):

Geometric mean (GM) is a statistical measure of central tendency, useful for multiplicative averaging and growth rates, ratios, and multiplicative relationships, unlike the arithmetic mean, which considers the product of values.

Properties of Geometric Mean:

- Applicable only to positive numbers.
- Influenced by outliers, stronger than the arithmetic mean due to multiplicative nature.
- Order Sensitivity: Changes with multiplication or division by a constant.
- Useful in multiplicative processes like growth rates, compound interest, and ratios.

Mean as a Measure of Central tendency

Geometric Mean for Ungrouped Data:

The formula for the geometric mean of n numbers $x_1, x_2, \dots, x_n$ is:

$$Geometric\ Mean = (x_1 \times x_2 \times \dots \times x_n)^{\frac{1}{n}} = \prod_{i=1}^{n} x_i^{\frac{1}{n}}$$

Where,

- $\prod_{i=1}^{n} x_i$ denotes the product of all n values
- n is the total number of values

Geometric Mean for Grouped Data: To calculate the geometric mean for grouped data, you need to follow these steps:

- Identify the midpoint of each class interval.
- Calculate the relative frequency for each class interval.

Mean as a Measure of Central tendency

Geometric Mean for Grouped Data:

- Raise each midpoint to the power of its relative frequency.
- Multiply all the results obtained from step 3.
- Take the nth root of the product, where n is the total number of data points.
- The formula for geometric mean for grouped data is,

$$Geometric\ Mean = \left(x_1^{f_1} \times x_2^{f_2} \times \ldots \times x_n^{f_n}\right)^{\frac{1}{n}} = \prod_{i=1}^{n} {x_i^{f_i}}^{\frac{1}{n}}$$

Where:

- x_i = midpoint of the class interval i
- f_i = relative frequency of class interval i
- n = total number of data points

Mean as a Measure of Central tendency

Example (Ungrouped Data):

Suppose we want to calculate the geometric mean of the scores of 5 students in a math test:

85	90	92	88	95

To calculate the geometric mean for this ungrouped data, you would follow these steps:

- Multiply all the numbers together: $85 \times 90 \times 92 \times 88 \times 95 = 37524000$
- Since there are 5 numbers, take the 5th root of the product:
- $\sqrt[5]{37524000} \approx 90.02$

So, the geometric mean of the scores is approximately 90.02.

Mean as a Measure of Central tendency

Example (Grouped Data):

Now, let's consider a scenario where we have grouped data. Suppose we want to calculate the geometric mean of the test scores of 40 students in a class, but the scores are grouped into intervals. Here's the frequency distribution:

Score Interval	Mid-point (x_i)	Frequency (f_i)	Relative Frequency (f_i)	$x_i^{f_i}$
80	82.5	5	0.125	1.736
85	87.5	12	0.30	3.825
90	92.5	15	0.375	5.461
95	97.5	8	0.20	2.499

Mean as a Measure of Central tendency

To calculate the geometric mean for this grouped data, you would follow the steps outlined in the previous response:

$$\prod_{i=1}^{n} {x_i}^{f_i}{}^{\frac{1}{n}} = (1.736 \times 3.825 \times 5.461 \times 2.499)^{\frac{1}{40}} = 1.119$$

So, the geometric mean of grouped data is 1.119.

Harmonic Mean:

The harmonic mean is a mathematical and statistical average used in rates, ratios, and other situations involving multiple quantities. It is calculated by dividing the number of observations by the sum of their reciprocals.

Harmonic Mean (HM) for Ungrouped Data:

Mathematically, the harmonic mean (H) of n observations $x_1, x_2, \dots, x_n$ is represented as:

$$H = \frac{n}{1/x_1 + 1/x_2 + \dots + 1/x_n}$$

Mean as a Measure of Central tendency

Example (Ungrouped Data):

Suppose a primary school teacher wants to calculate the average speed of students during a running race. She records the speeds (in meters per second) of 5 students as follows: 2, 3, 4, 5, and 6. To find the harmonic mean of these speeds, we first find the reciprocals of each speed:

Reciprocal $$H = \frac{5}{\frac{1}{2}+\frac{1}{3}+\frac{1}{4}+\frac{1}{5}+\frac{1}{6}}$$

Then, we find the arithmetic mean of these reciprocals:

$$H = \frac{5}{\frac{263}{360}}$$

$$H = \frac{1800}{263}$$

$$H = 6.84\ ms^{-1}.$$

Mean as a Measure of Central tendency

Example (Ungrouped Data):

$$H = \frac{5}{\frac{1}{2}+\frac{1}{3}+\frac{1}{4}+\frac{1}{5}+\frac{1}{6}}$$

$$H = \frac{5}{\frac{263}{360}}$$

$$H = \frac{1800}{263}$$

$$H = 6.84\ ms^{-1}.$$

So, in this example, the harmonic mean of the speeds of the students is approximately 6.84 meters per second.

Mean as a Measure of Central tendency

Harmonic Mean for Grouped Data: For grouped data, the formula for calculating the harmonic mean is adjusted to accommodate the intervals or classes in the data.

Let's assume we have grouped data with intervals $x_1, x_2, \ldots, x_n$ and corresponding frequencies $f_1, f_2, \ldots, f_n$. The formula is:

$$\textit{Harmonic Mean} = \frac{n}{\frac{f_1}{x_1} + \frac{f_2}{x_2} + \ldots + \frac{f_n}{x_n}} = \frac{n}{\sum_{i=1}^{k} \frac{f_i}{x_i}}$$

where,

- x_i = midpoint of the class interval i
- f_i = relative frequency of class interval i
- n = total number of data points
- k is the number of intervals.

Mean as a Measure of Central tendency

Example (Grouped Data): A primary school teacher calculates average homework time by grouping data into time intervals and recording the number of students in each interval.

Time Interval (hours)	Mid-value (Frequency)	Number of Students (Frequency)
0 - 1	0.5	5
1 - 2	1.5	8
2 - 3	2.5	4
3 - 4	3.5	2
4 - 5	4.5	1

To find the harmonic mean of the time spent on homework, the teacher can use the formula:

$$\textit{Harmonic Mean} = \frac{n}{\frac{f_1}{x_1} + \frac{f_2}{x_2} + \ldots + \frac{f_n}{x_n}} = \frac{n}{\sum_{i=1}^{k} \frac{f_i}{x_i}}$$

Mean as a Measure of Central tendency

Here, $n = 5 + 8 + 4 + 2 + 1 = 20$

$$HM = \frac{20}{\frac{5}{.5} + \frac{8}{1.5} + \frac{4}{2.5} + \frac{2}{3.5} + \frac{1}{4.5}}$$

$$HM = \frac{20}{10 + 2.133 + 3.2 + .571 + .222}$$

$$H = \frac{20}{16.126}$$

$$H = \frac{3}{0.1567}$$

$$H \approx 1.239$$

So, the harmonic mean of the time spent on homework by the students in this class is approximately 1.239 hours.

Mean as a Measure of Central tendency

Weighted Mean: The weighted mean is an average that assigns different weights to different values in a dataset, calculated by multiplying each value by its weight, summing, and dividing.

Weighted Mean (WM) for Ungrouped Data:

Mathematically, the formula for the weighted mean of a dataset with n values, each with its corresponding weight

$w_1, w_2, \ldots, w_n$: is

$$Weighted\ Mean\ (WM) = \frac{w_1 \times x_1 + w_2 \times x_2 + \ldots + w_n \times x_n}{w_1 + w_2 + \ldots + w_n}$$

Example (Ungrouped Data):

Let's consider an example where we want to calculate the weighted mean of students' test scores. Here's a set of ungrouped data representing the test scores of 5 students:

Mean as a Measure of Central tendency

85	90	92	88	85

Now, let's say we have additional information that certain students are considered more representative of the overall performance of the class due to their consistent performance or other factors. We assign weights to each student accordingly:

Test Score	85	90	92	88	85
Weight	1	2	1	2	1

$$Weighted\ Mean\ (WM) = \frac{(85 \times 1) + (90 \times 2) + (92 \times 1) + (88 \times 2) + (85 \times 1)}{1 + 2 + 1 + 2 + 1}$$

$$= \frac{618}{7}$$

$$= 88.285$$

So, the weighted mean of the student's test scores for this primary school is 88.285.

Mean as a Measure of Central tendency

Weighted Mean (WM) for Grouped Data:

Grouped data involves values grouped into categories, each with a weight assigned to it. The grouped weighted mean is calculated by calculating the weighted mean of these groups.

Suppose you have grouped data with intervals $x_1, x_2, \dots, x_k$ and corresponding frequencies $f_1, f_2, \dots, f_k$, and each group has its weight $w_1, w_2, \dots, w_k$. The formula for calculating the grouped weighted mean is:

$$Weighted\ Mean\ (WM) = \frac{w_1 \times \overline{x_1} + w_2 \times \overline{x_2} + \dots + w_n \times \overline{x_n}}{w_1 + w_2 + \dots + w_n}$$

Example (Grouped Data):

A teacher could group the test scores into intervals (e.g., 0-50, 51-75, 76-100) and calculate the average scores for each group. Then, each group's average score would be weighted according to its importance.

Mean as a Measure of Central tendency

Group interval	Mean ($\bar{x}_i$)	Weights
0 - 50	40	1
51 – 75	65	2
76 - 100	90	1

Total number of students = 20 (from Class A) + 15 (from Class B) + 25 (from Class C) = 60 students

Now, let's calculate the weights:

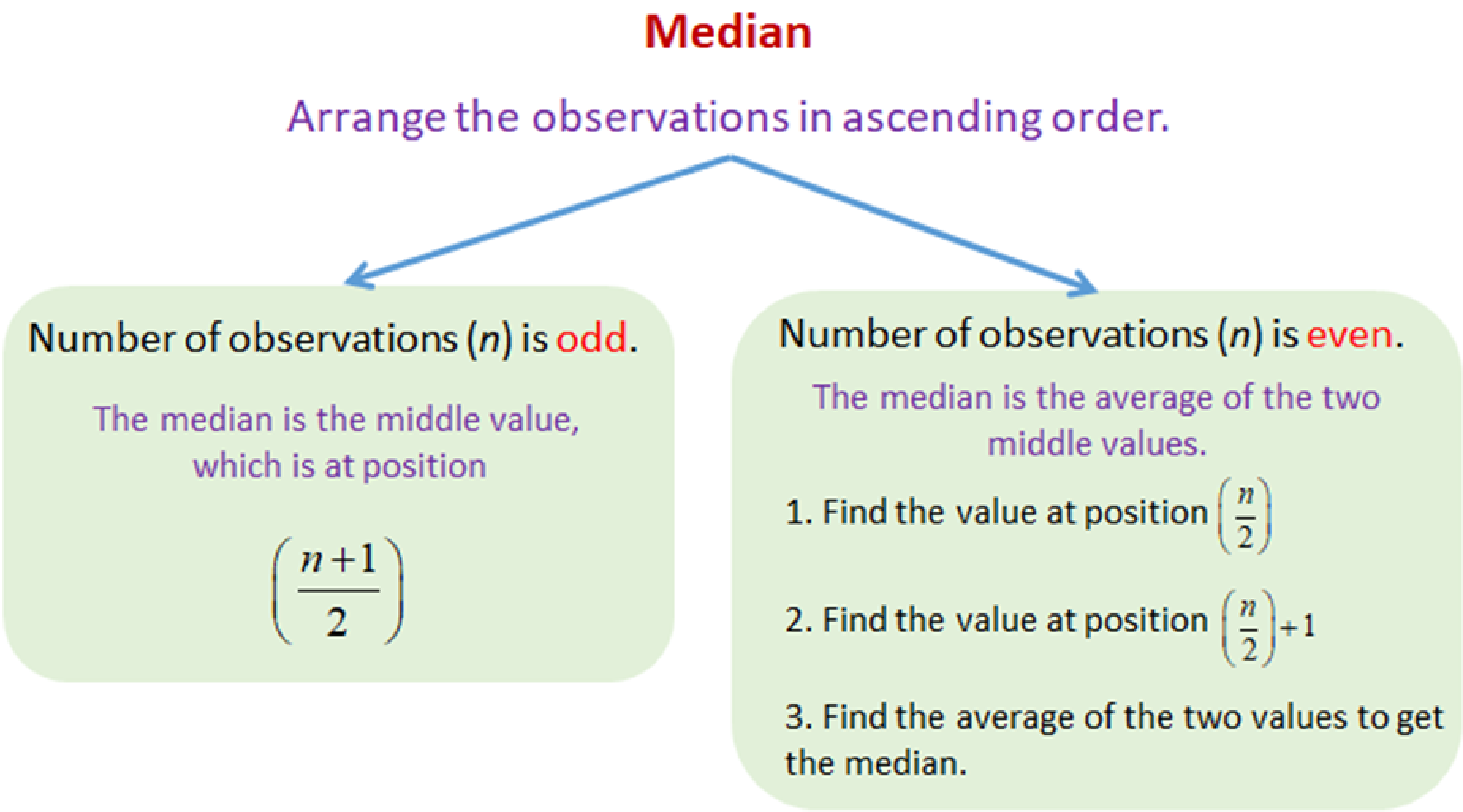
$$Weighted\ Mean\ (WM) = \frac{1\times 40 + 2\times 62 + 90\times 1}{1+2+1} = \frac{254}{4} = 63.5$$

So, the weighted mean of the students' test scores for this primary school is approximately 63.4

Median as a Measure of Central tendency

The median is a statistical measure used to find the middle value of a dataset, often used as a central tendency due to its sensitivity to extreme values. To calculate the median:

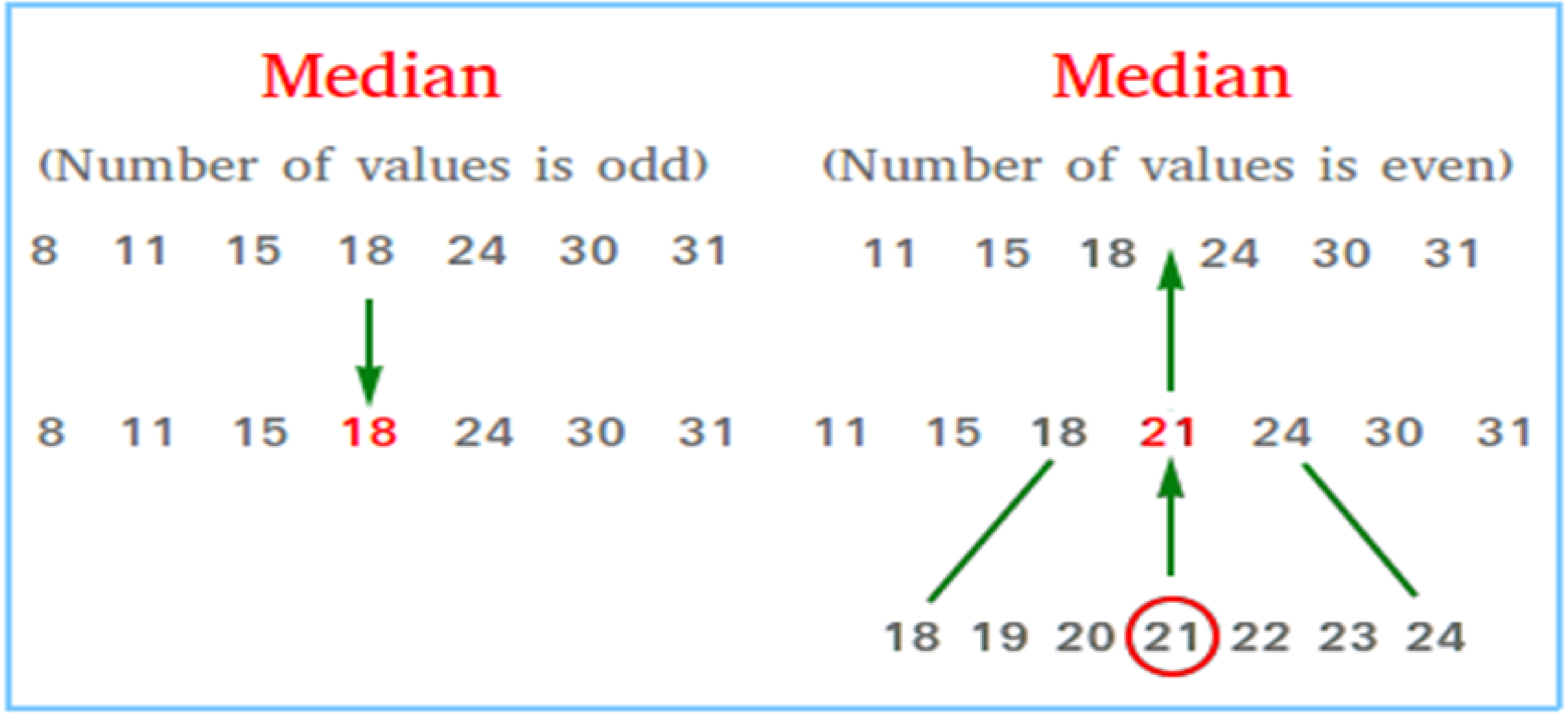
- Arrange the dataset in ascending or descending order.
- If the dataset has an odd number of values, the median is the middle value.
- If the dataset has an even number of values, the median is the average of the two middle values.



Median as a Measure of Central tendency

Example:

- If the dataset has an odd number of values, the median is simply the middle value. For example, in the dataset {8,11,15,18,24,30,31}, the median is 18.
- If the dataset has an even number of values, the median is the average of the two middle values. For example, in the dataset {11,15,18,24,30,31}, the two middle values are 18 and 24, so the median is (18+24)/2=21.



Importance of Median

- **Grading and Assessment:** Uses median scores to determine central tendency of student performance.
- **Resource Allocation:** Provides fair representation of typical situations in resource allocation.
- **Classroom Placement:** Uses median performance level as a benchmark for class placement decisions.
- **Evaluating Trends:** Helps identify trends and patterns in student achievement.
- **Benchmarking and Comparisons:** Uses median scores as benchmarks for comparison with other schools or districts.
- **Addressing Achievement Gaps:** Identifies achievement gaps and disparities in educational outcomes.

Median as a Measure of Central tendency

Median for Grouped Data

To find the median in grouped data, identify the class interval containing the median value and use interpolation to determine the exact median within that interval.

Calculate the Median:

$$Median = L + \left(\frac{\frac{N}{2} - C}{f}\right) \times w$$

Where,

- L is the lower boundary of the median class interval,
- N is the total number of observations,
- C is the cumulative frequency of the class interval preceding the median class interval,
- *f* is the frequency of the median class interval,
- w is the width of the median class interval.

Median as a Measure of Central tendency

Example: Suppose we have a grouped data representing test scores

Solution:

Class Interval	Frequency	Cumulative Frequency
60-69	5	5
70-79	8	13
80-89	12	25
90-99	10	35
100-109	5	40
Total	40	

Calculate the Median:

$$L = 80\,,\ N = 40\,,\ C = 13\,,\ f = 12\,,\ w = 10$$

$$Median = 80 + \left(\frac{\frac{40}{2} - 13}{12}\right) \times 10 = 80 + 5.833 \approx 85.833$$

So, the median test score for this grouped data is approximately 85.833.

Mode as a Measure of Central tendency

The mode is a statistical measure that represents the most frequently occurring values in a dataset, which can be one (unimodal), two (bimodal), or more than two (multimodal).

Example: Tracking the favorite colors of students in a classroom. Let's say a teacher surveyed a class of 20 students and asked them to name their favorite color. After collecting their responses, he might find that:

Six students preferred **blue**, eight preferred **red**, three preferred **green**, and three preferred **yellow**.

In this scenario, the mode, or the most frequently occurring favorite color, is red because it was chosen by the highest number of students.

Mode as a Measure of Central tendency

Mode for Grouped Data

Finding the mode for grouped data involves identifying the class interval with the highest frequency, and then using interpolation to estimate the modal value within that interval. Here's how you can find the mode for grouped data:

Calculate the Mode:

Use the formula for modal value (mode) within a class interval:

$$Mode = L + \left(\frac{f_m - f_{m-1}}{2f_m - f_{m-1} - f_{m+1}}\right) \times w$$

Where,

- L is the lower boundary of the modal class interval,
- f_m is the frequency of the modal class interval,
- f_{m-1} is the frequency of the class interval before the modal class intervals (if any),
- f_{m+1} is the frequency of the class interval after the modal class interval (if any),
- w is the width of the modal class interval.

Mode as a Measure of Central tendency

Example (Grouped Data):

Suppose we have the following grouped data representing test scores:

Class Interval	Frequency
60-69	5
70-79	8
80-89	12
90-99	10
100-109	5

Solution: Calculate the Mode

Using the formula: $L = 80$, $f_m = 12$, $f_{m-1} = 8$, $f_{m+1} = 10$, $w = 10$

$$Mode = L + \left(\frac{f_m - f_{m-1}}{2f_m - f_{m-1} - f_{m+1}}\right) \times w$$

$$Mode = 80 + \left(\frac{12 - 8}{2 * 12 - 8 - 10}\right) \times 12 = 80 + 6.667 \approx 86.667$$

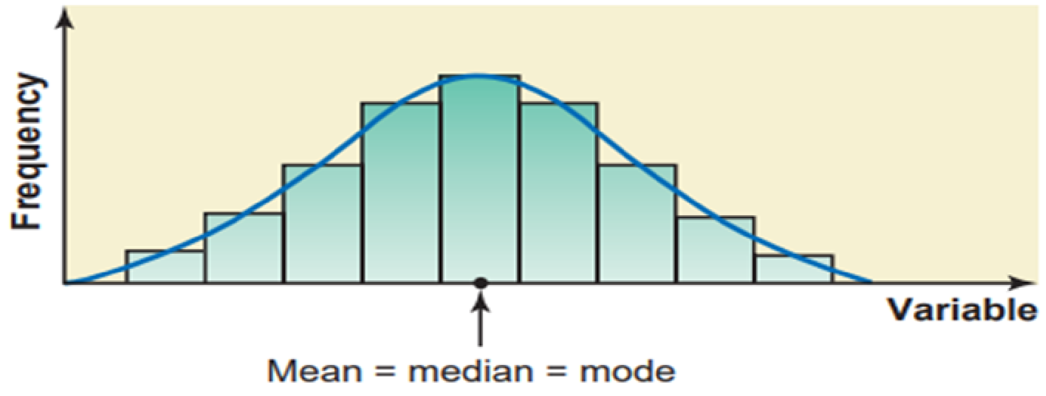
So, the mode for this grouped data is approximately 86.667.

Importance of Mode

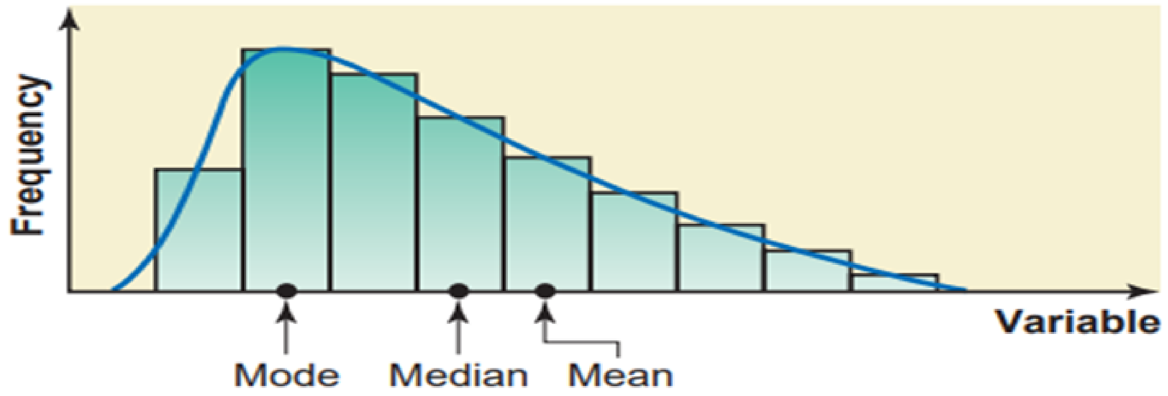
- Understanding Student Preferences: Informs teaching methodologies and curriculum design.
- Identifying Common Challenges: Identifies areas of struggle.
- Resource Allocation: Helps allocate resources effectively.
- Curriculum Development: Informs curriculum development and course offerings.
- Assessment and Evaluation: Provides insights into student performance distribution

Empirical Relationship among Mean, Median and Mode

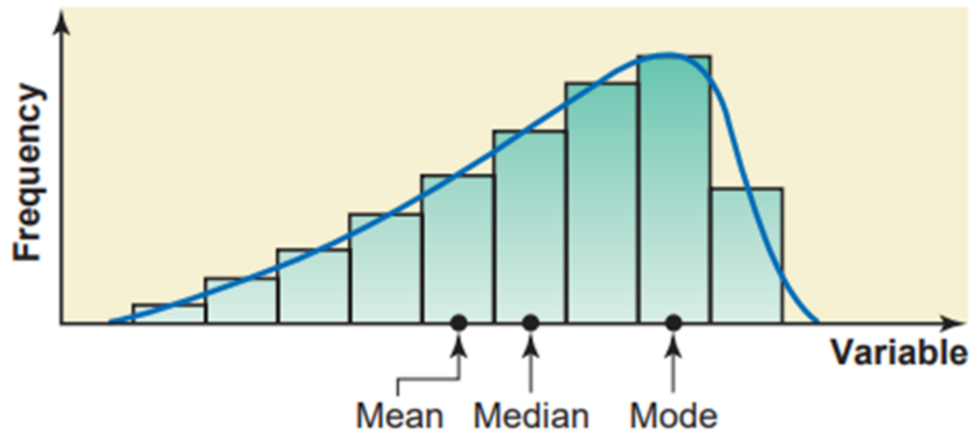
1. For a symmetric histogram and frequency distribution



2. For a histogram and a frequency distribution curve skewed to the right:



3. For a histogram and a frequency distribution curve skewed to the left:



Mean

Advantages

- Sensitive to All Data Points
- Mathematically Convenient
- Balancing Effect

Disadvantages

- Sensitive to Outliers
- Not Robust to Skewed Distributions
- Not Suitable for Categorical Data

Median

Advantages

- Robust to Outliers
- Suitable for Ordinal Data
- Clear Interpretation

Disadvantages

- Not Sensitive to All Data Points
- Limited Mathematical Properties
- More Complex for Grouped Data

Mode

Advantages

- Applicable to Nominal Data
- Easy to Identify
- Robust to Outliers

Disadvantages

- Not Unique
- May Not Represent Central Tendency
- Limited Applicability

Thank You