Measures of Central Tendency and Measures of Location

Munmun Akter

Lecturer

NFE, DIU

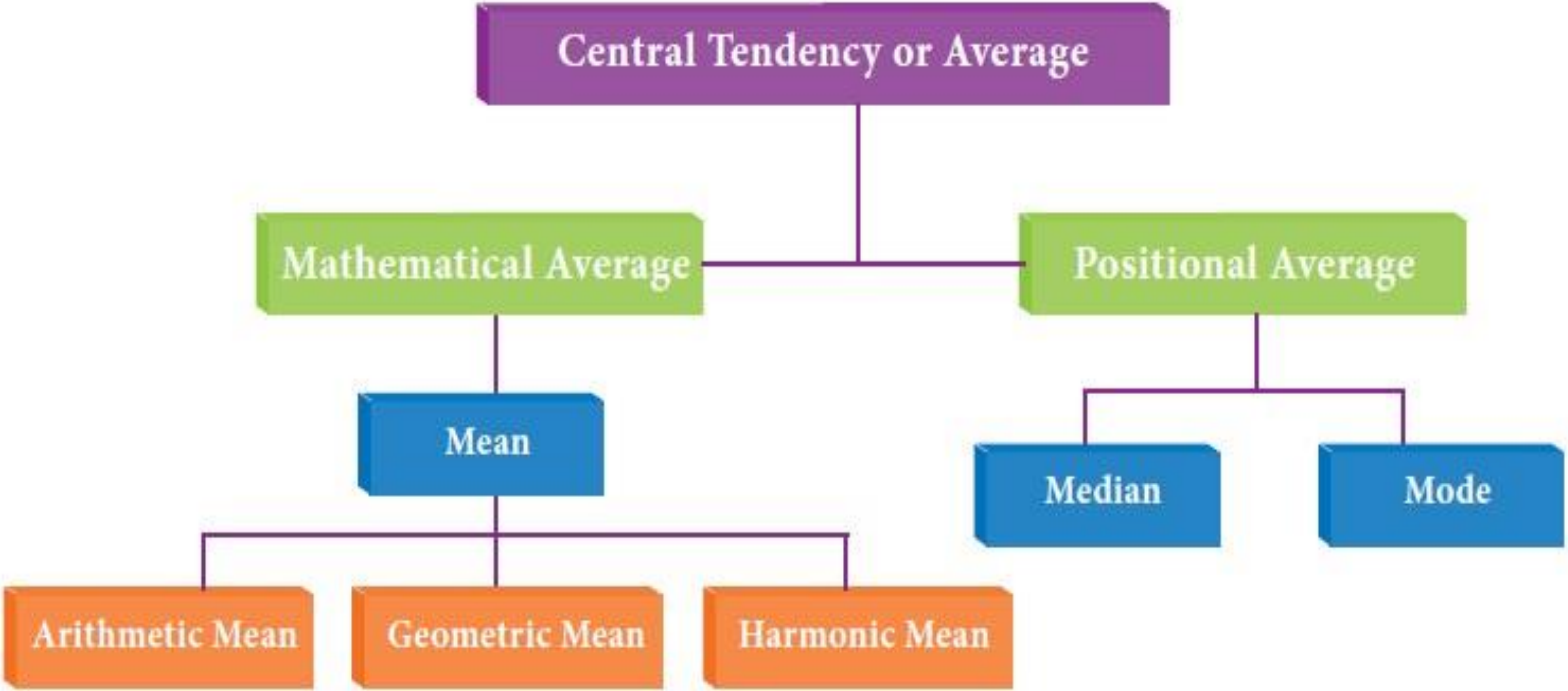
Central tendency

- The central tendency is stated as the statistical measure that represents the single value of the entire distribution or a dataset.
- It helps to summarize and understand the main characteristic of a distribution or dataset.
- Three most common measures of central tendency:

1. the Mean,
2. the Median, and
3. the Mode.

- Each measure of central tendency has its advantages and is appropriate for different situations.
- The choice of which measure to use depends on the nature of the data and the purpose of the analysis.

Mean or Average



Arithmetic Mean/ Mean

- The most common measure of central tendency.
- It is simply the sum of the numbers divided by the number of numbers.
- The symbol "μ" is used for the mean of a population.
- The symbol "M" is used for the mean of a sample.
- Mathematically, if you have a set of numbers {x1, x2, x3, ..., xn}, the arithmetic mean (usually denoted by "μ" or "x̄") is,

  Mean = (x1 + x2 + x3 + ... + xn) / n

  Arithmetic Mean for Grouped Data = (Σ (m * f)) / N

m= mid point, f=frequency, n= total number

Advantages 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 readily put to algebraic treatment.
- It is least affected by fluctuations of sampling.

Disadvantages of Arithmetic mean

- The arithmetic mean is highly affected by extreme values.
- It cannot average the ratios and percentages properly.
- It cannot be computed accurately if any item is missing.
- The mean sometimes does not coincide with any of the observed value.
- It cannot be computed accurately for unequal and open-ended class intervals.

Geometric Mean (GM)

- Is a type of average, usually used for growth rates, like population growth or interest rates. While the arithmetic mean **adds** items, the geometric mean **multiplies** items.
- Also, you can only get the geometric mean for positive numbers.
- Mathematically, if you have a set of positive numbers {x1, x2, x3, ..., xn}, the geometric mean (GM) is calculated as:

  GM = (x1 * x2 * x3 * ... * xn) ^ (1/n)

- Consider the dataset of positive numbers: {2, 4, 8, 16} for calculating GM

Harmonic Mean (HM)

- It is used to calculate the central tendency of a set of positive numbers.
- It is especially useful when dealing with rates, ratios, and other situations where the reciprocal of values is important.

- To compute the harmonic mean of a set of positive numbers, you take the reciprocal of each number, calculate the arithmetic mean of these reciprocals, and then take the reciprocal of the result.

Mathematically, if you have a set of positive numbers $\{x1, x2, x3, ..., xn\}$, the harmonic mean (HM) is calculated as:

$$HM = n / \{(1/x1) + (1/x2) + (1/x3) + ... + (1/xn)\}$$

Or equivalently:

$$HM = n / (\sum(1/xi) \text{ from } i=1 \text{ to } n)$$

Now consider the dataset of positive numbers: {2, 4, 8, 16} for calculating HM.

Weighted Mean (WM)

- is a type of average that takes into account the importance or significance of each data point in a dataset.
- Unlike the simple arithmetic mean, where each value contributes equally to the average.
- It assigns specific weights to each value based on its relative importance.

Mathematically, if you have a set of values {x1, x2, x3, ..., xn} with corresponding weights {w1, w2, w3, ..., wn}, the weighted mean (WM) is calculated as:

WM = (w1 * x1 + w2 * x2 + w3 * x3 + ... + wn * xn) / (w1 + w2 + w3 + ... + wn)

In this formula, the weight (wi) represents the relative importance or significance of each data point (xi).

The sum of all the weights is typically equal to 1, although it can be any positive value.

Now consider the dataset: {10, 20, 30, 40} and the corresponding weights: {0.1, 0.2, 0.3, 0.4} for calculating WM

Median

- It represents the middle value in a dataset when the data points are arranged in ascending or descending order.
- It is useful when dealing with datasets that may have extreme values (outliers) that can significantly affect the mean.
- By taking the middle value, the median is less influenced by outliers and provides a more robust representation of the "typical" value in the dataset.

To find the median of a dataset, follow these steps:
1. Arrange the data in ascending or descending order.
2. If the dataset has an odd number of observations, the median is the middle value.
3. If the dataset has an even number of observations, the median is the average of the two middle values.

Consider the dataset: {10, 20, 25, 30, 40} for median calculation.

$$Median = L + \left(\frac{N}{2} - C.F.\right)\frac{h}{f}$$

Where;

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

$N = total\ frequency$

$C.F. = cumulative\ frequency(less\ than\ type)$
$of\ the\ class\ preceding\ the\ median\ class$

$f = frequency\ of\ median\ class$

$h = width\ of\ median\ class$

Marks	Number of students (f)	Cumulative frequency (cf)
0 - 10	2	2
10 - 20	7	9
20 - 30	15	24
30 - 40	10	34
40 - 50	11	45
50 - 60	5	50
	N = 50	

Mode

- It represents the value that appears most frequently in a dataset. In other words, it is the data point with the highest frequency.
- To find the mode of a dataset, you simply identify the value that occurs most often. A dataset can have one mode (unimodal) if there is one value that appears more frequently than any other.
- It can also have two modes (bimodal) if two values have the same highest frequency, or more than two modes (multimodal) if multiple values share the highest frequency.

- The mode is particularly useful when you want to identify the most common or typical value in a dataset.
- It is not influenced by extreme values or the overall spread of data, making it helpful in datasets that may have outliers or a skewed distribution.

Consider the dataset: {10, 20, 30, 20, 40, 50, 30, 30} for calculating mode.

$$M_o = l + \left(\frac{f_1 - f_0}{2f_1 - f_0 - f_2}\right)h$$

Where

l = lower limit of the modal class,

h = size of the class interval (assuming all class sizes to be equal),

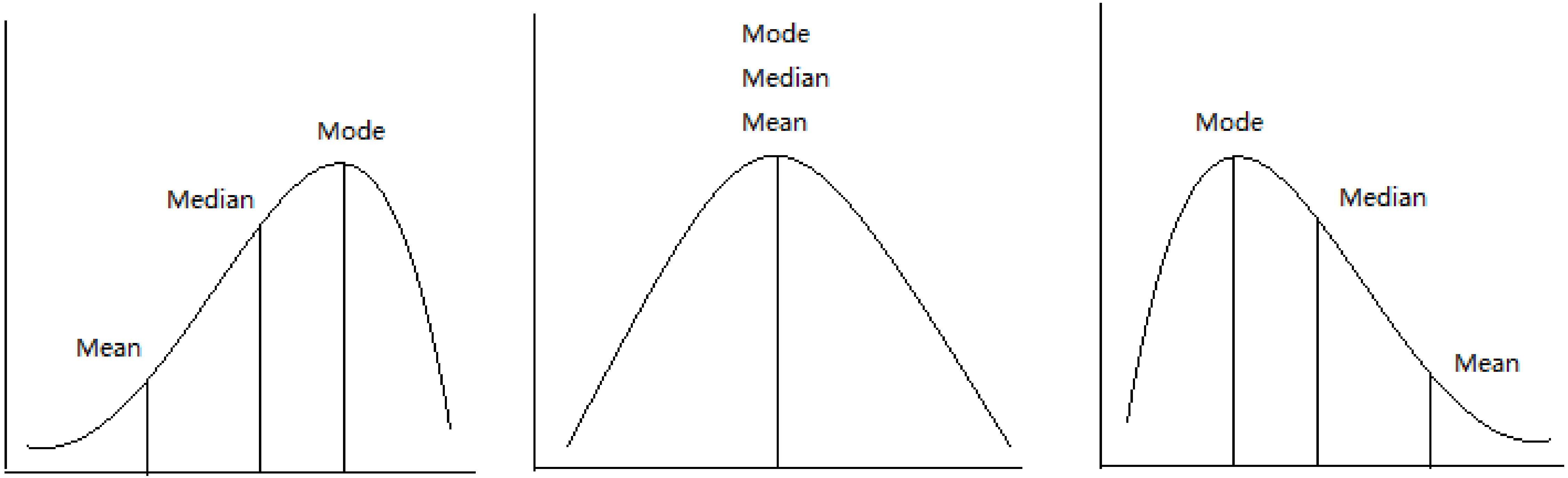
f_1 = frequency of the modal class,

f_0 = frequency of the class preceding the modal class,

f_2 = frequency of the class succeeding the modal class.

Exam Score	Frequency
51-60	4
61-70	8
71-80	15
81-90	8
91-100	5

Central Tendency Measure	Pros	Cons
Mean	Sensitive as it takes all data values into account(reliable)	Biased output if outliers/extreme values exist in the data set
Median	Not affected by extreme values	-Less sensitive than Mean as it only focusses on giving out the middle data point irrespective of how far the other values are from the middle -Needs the data to be arranged in the ascending order before computing
Mode	Not affected by extreme values and can be used with non-numerical data	There may be more than one mode or no mode at all and it may not reflect data summary accurately



Left skew Normal Distribution Right skew

Thank You

Any Question?