UNIT 3

Measures of Location/ Central Tendency

A measure of central tendency/ Location

A measure of central tendency is a single value that attempts to describe a set of data by identifying the central position within that set of data. As such, measures of central tendency are sometimes called **measures of central location**. They are also classed as summary statistics. The mean (often called the average) is most likely the measure of central tendency that you are most familiar with, but there are others, such as the **median and the mode.**

The **mean, median and mode** are all valid measures of central tendency, but under different conditions, some measures of central tendency become more appropriate to use than others.

Mean

The mean is the arithmetic average, and it is probably the measure of central tendency that you are most familiar. Calculating the mean is very simple. You just add up all of the values and divide by the number of observations in your dataset.

The three classical Pythagorean means are

- The arithmetic mean(AM)
- The geometric mean(GM), and
- The harmonic mean(HM).

The Arithmetic Mean:

The arithmetic mean is calculated by adding all of the numbers and dividing it by the total number of observations in the dataset.

For example: Arithmetic Mean of 4 + 10 + 7 is 21/3 = 7

For raw data $\textbf{Arithmetic Mean} = \frac{\sum \boldsymbol{x}}{\boldsymbol{n}}$**,** where $\sum x$ is the sum of all individual's data and ***n*** is the total number of data/observation.

For frequency distribution $\textbf{Arithmetic Mean A. M.} = \frac{\sum fx}{\sum f}, where\ f\ is\ the\ frequeny$

For Example:

For Ungrouped Data

x	5	10	15	20	25	30
f	4	5	7	4	3	2

x	f	f x x = fx
5	4	20 (4x5)
10	5	50 (5x10)
15	7	105
20	4	80
25	3	75
30	2	60
Total	**N=25**	**Σfx =390**

$$\overline{\boldsymbol{x}} = \frac{\sum fx}{\sum f} = \frac{390}{25} = 15.6$$

For Grouped Data

Class interval	f1	Class mark(x1)	f1x1
0-5	4	2.5	10
5-10	6	7.5	45
10-15	10	12.5	125
15-20	16	17.5	280
20-25	12	22.5	270
25-30	8	27.5	220
30-35	4	32.5	130
TOTAL	**$\sum$ f1=60**		**$\sum$ f1x1=1080**

$$\overline{\boldsymbol{x}} = \frac{\sum fx}{\sum f} = \frac{1080}{60} = 18$$

Note

The arithmetic mean works well when the data is in an additive relationship between the numbers, often when the data is in a ‘linear’ relationship which when graphed the numbers either fall on or around a straight line. *i.e*. when they are **clustered.**

Geometric Mean

Not all datasets establish a linear relationship, sometimes you might expect a multiplicative or exponential relationship and, in those cases, arithmetic mean is ill-suited and might be misleading to summarize the data.

The Geometric Mean (GM) is the average value or mean which signifies the central tendency of the set of numbers by taking the root of the product of their values. Basically, we multiply the '**n**' values altogether and take out the **nth root** of the numbers, where n is the total number of values.

$\boldsymbol{Geometric\ Mean}\quad G.M. = \sqrt[n]{x_1\ x_2\ \ x_3\ .\ .\ .\ .\ x_n}$

i.e. The geometric mean of 5, 7 and 10 is $= \sqrt[3]{5 \times 7 \times 10} = 7.04$

Note

The geometric mean works well when the data is in an multiplicative relationship or in cases where the data is compounded; hence you multiply the numbers rather than add all the numbers.

For example

Suppose you invested $500 initially which yielded 10% return the first year, 20% return the second year and 30% return the third year. After three years, you have $500 * 1.1 * 1.2 * 1.3 = $858.00.

Whereas if you taking arithmetic mean, it's 10+20+30 = 20% return on average per year, so after three years you would have $500 * 1.2 * 1.2 * 1.2 = $864. As we can see, arithmetic mean overestimates earnings by nearly $6 which is not right since we applied an additive operation to a multiplicative process.

Investors usually consider using geometric mean over arithmetic mean to measure the performance of an investment or portfolio.

Harmonic Mean

The Harmonic Mean (HM) is defined as the reciprocal of the arithmetic mean of the reciprocals of the data values.

i.e. $H.M. = \frac{1}{\left(\frac{1}{x_1}+\frac{1}{x_2}+\frac{1}{x_3}+---+\frac{1}{x_n}\right)/n} = \frac{n}{\frac{1}{x_1}+\frac{1}{x_2}+\frac{1}{x_3}+---+\frac{1}{x_n}}$

For the numbers 4, 6 and 8 $H.M. = \frac{1}{\left(\frac{1}{4}+\frac{1}{6}+\frac{1}{8}\right)/3} = \frac{3}{\left(\frac{1}{4}+\frac{1}{6}+\frac{1}{8}\right)} = 5.54$

Note

Harmonic mean is used when we want to average units such as speed, rates and ratios.

For example: I drove at an speed of 60km/hr to Seattle downtown and returned home at a speed of 30km/hr and the distance from my house to Seattle is 20 km. What was my average speed for the whole trip?

Average speed = $\frac{1}{\left(\frac{1}{60}+\frac{1}{30}\right)/2} = 40\ km/h$ **NOT** (60+30)/2 = 45 km/hr.

Relationship among AM, GM and HM

For two Number a and b

Arithmetic mean (A.M.) = $\frac{a+b}{2}$

Geometric mean (G.M.) = $\sqrt{a \times b}$

Harmonic mean (H.M.) = $\frac{1}{(\frac{1}{a}+\frac{1}{b})/2} = \frac{2ab}{a+b} = \frac{ab}{(a+b)/2} = \frac{(GM)^2}{AM}$

$$HM = \frac{(GM)^2}{AM}$$

The harmonic mean has the least value compared to the geometric and arithmetic mean and $AM \geq GM \geq HM$

Median

Like mean median is a measure of central tendency. Median determines the middle value of a dataset listed in ascending order (i.e., from smallest to largest value). The measure divides the lower half from the higher half of the dataset.

How to Find the Median

The median can be easily found. In some cases, it does not require any calculations at all. The general steps of finding the median include:

- Arrange the data in ascending order (from the lowest to the largest value).
- Determine whether there is an even or an odd number of values in the dataset.
- If the dataset contains an odd number of values, the median is a central value that will split the dataset into halves.
- If the dataset contains an even number of values, find the two central values that split the dataset into halves. Then, calculate the mean of the two central values. That mean is the median of the dataset.

For Example

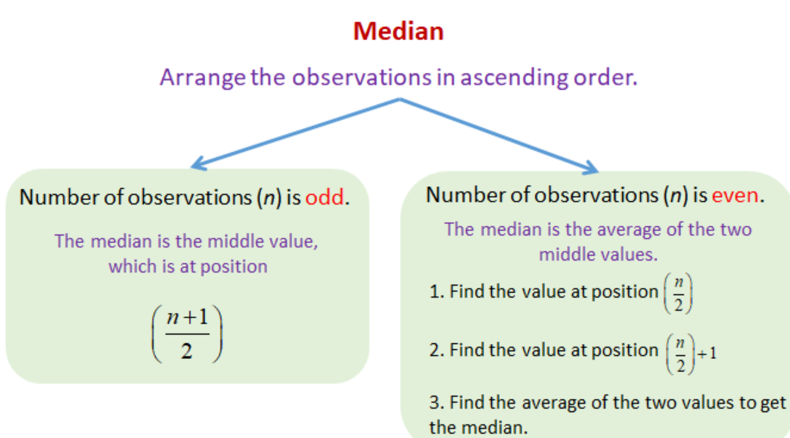
1, 3, 3, **6**, 7, 8, 9

Median = 6

1, 2, 3, **4**, **5**, 6, 8, 9

Median = (4 + 5) ÷ 2

= 4.5



Median Class

To find the median class, we have to find the cumulative frequencies of all the classes and n/2. After that, locate the class whose cumulative frequency is greater than (nearest to) n/2. The class is called the median class.

data :

Class	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60
Frequency	6	7	15	16	4	2

Class	Frequency	Cumulative frequency	Mid-point x_i
0 – 10	6	6	5
10 – 20	7	7 + 6 = 13	15
20 – 30	15	13 + 15 = 28	25
30 – 40	16	28 + 16 = 44	35
40 – 50	4	44 + 4 = 48	45
50 – 60	2	48 + 2 = 50	55
	50		

$N = \sum f_i = 50$

Median Class = $\left(\frac{N}{2}\right)^{th}$ term

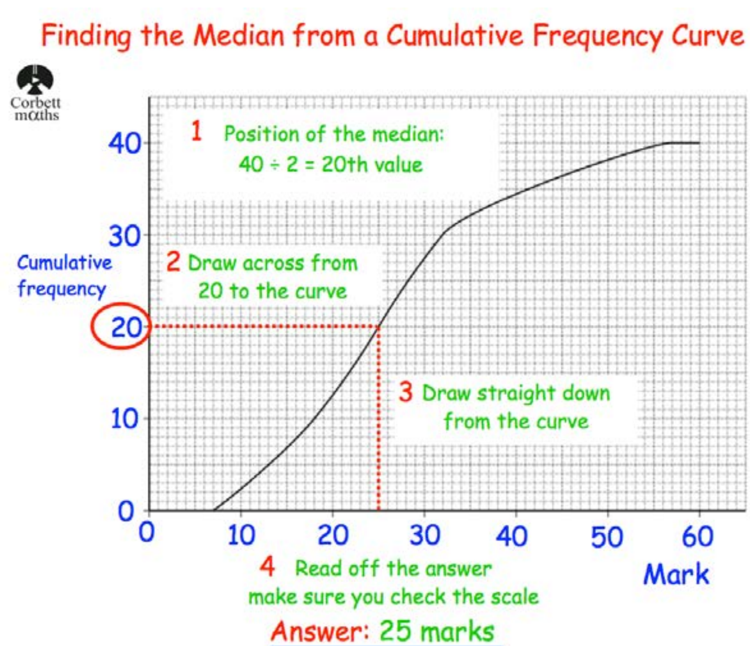
$= \left(\frac{50}{2}\right)^{th}$ term

$= 25^{th}$ term

In above data, cumulative frequency of class 20 - 30 is 28 which is slightly greater than 25.

∴ Median class = 20 - 30

Finding Median Using Cumulative Frequency Graph



Note

As Median does not get influenced by extreme values (mean does get influenced by extreme value), so when dataset is highly fluctuating or deviating from the central value, median can be used as an appropriate measure of central tendency.

Mode:

The mode is the value that appears most frequently in a data set. A set of data may have one mode, more than one mode, or no mode at all.

When the data set has one mode, we call it **Unimodal**

For example, the mode (unimodal) in the following dataset is 19:

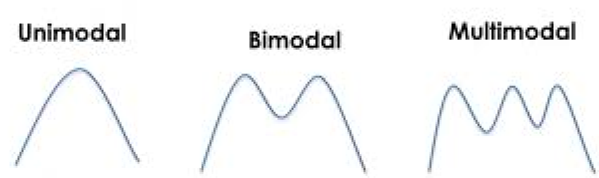
Dataset: 3, 4, 11, 15, 19, 19, 19, 22, 22, 23, 23, 26

When the data set has two modes, we call it **bimodal**

For example, the modes in the following dataset are 11 and 19:

Dataset: 3, 7, 4, 11, 15, 11, 14 19, 19, 19, 22, 20, 11, 22, 23, 23, 26

When the data set has more than two modes, we call it **multi-modal**



Note

The mode tells us the most common value in categorical data when the mean and median can't be used.