BASIC OF STATISTICS
UNIT-II
MEAN, MEDIAN AND MODE

DR. PAVAN KUMAR SINGH
ASSISTANT PROFESSOR
DEPARTMENT OF PHYSICS,
Faculty of Engineering,
Teerthanker Mahaveer University,
Moradabad

Measures of Central Tendency

- A *measure of central tendency* is a descriptive statistic that describes the average, or typical value of a set of scores.
- There are three common measures of central tendency:
  - Mean
  - Median
  - Mode

Arithmetic mean or mean

- Mean is an arithmetic average of the data set and it can be calculated by dividing a sum of all the data points with the number of data points in the data set.

$$Mean = \frac{Sum\ of\ observations}{Number\ of\ observations}$$

Mean Formula For Ungrouped Data

The formula to find the mean of an ungrouped data is given below:

Suppose $x_1, x_2, x_3, \ldots, x_n$ be n observations of a data set, then the mean of these values is:

$$\bar{x} = \frac{\sum x_i}{n}$$

Here,

x_i = ith observation, $1 \le i \le n$

$\sum x_i$ = Sum of observations

n = Number of observations

Examples

Question 1: Find the mean of the following data set.

10, 20, 36, 12, 35, 40, 36, 30, 36, 40

Solution:

Given,

x_i = 10, 20, 36, 12, 35, 40, 36, 30, 36, 40

n = 10

Mean = $\sum x_i/n$

= (10 + 20 + 36 + 12 + 35 + 40 + 36 + 30 + 36 + 40)/10

= 295/10

= 29.5

Therefore, the mean of the given data set is 29.5.

Mean Formula For Grouped Data

- There are three methods to find the mean for grouped data, depending on the size of the data. They are:
- Direct Method
- Assumed Mean Method
- Step-deviation Method

Direct Method

Suppose $x_1, x_2, x_3, \ldots, x_n$ be n observations with respective frequencies $f_1, f_2, f_3, \ldots, f_n$. This means, the observation x_1 occurs f_1 times, x_2 occurs f_2 times, x_3 occurs f_3 times and so on. Hence, the formula to calculate the mean in the direct method is:

$$\bar{x} = \frac{f_1x_1+f_2x_2+f_3x_3+\ldots+f_nx_n}{f_1+f_2+f_3+\ldots+f_n}$$

Or

$$\bar{x} = \frac{\sum_{i=1}^{n} f_ix_i}{\sum_{i=1}^{n} f_i}$$

Here,

$\sum f_ix_i$ = Sum of all the observations

$\sum f_i$ = Sum of frequencies or observations

This method is used when the number of observations is small.

Example: Find the mean of the following distribution, which gives the scores obtained by the students in a quiz.

Marks	25	43	38	42	33	28	29	20
Number of students	20	1	4	2	15	24	28	6

Solution:

Let us create a table to find the sum:

Marks (x_i)	Number of students (f_i)	f_ix_i
25	20	500
43	1	43
38	4	152
42	2	84
33	15	495
28	24	672
29	28	812
20	6	120
Sum	100	2878

Mean = $(\sum f_ix_i)/\sum f_i$

= 2878/100

= 28.78

Thus, the mean of the given distribution is 28.78.

Assumed Mean Method

- In statistics, the assumed mean method is used to calculate mean or arithmetic mean of a grouped data. If the given data is large, then this method is recommended rather than a direct method for calculating mean. This method helps in reducing the calculations and results in small numerical values.

Assumed Mean Method Formula

Let $x_1, x_2, x_3, \ldots, x_n$ are mid-points or class marks of n class intervals and $f_1, f_2, f_3, \ldots, f_n$ are the respective frequencies. The formula of the assumed mean method is:

$$\bar{x} = a + \frac{\sum f_i d_i}{\sum f_i}$$

Here,

a = assumed mean

f_i = frequency of ith class

$d_i = x_i - a$ = deviation of ith class

Σf_i = n = Total number of observations

x_i = class mark = (upper class limit + lower class limit)/2

Example: The following table gives information about the marks obtained by 110 students in an examination.

Class	0-10	10-20	20-30	30-40	40-50
Frequency	12	28	32	25	13

Find the mean marks of the students using the assumed mean method.

Solution:

Class (CI)	Frequency (f_i)	Class mark (x_i)	$d_i = x_i - a$	$f_i d_i$
0-10	12	5	5 − 25 = − 20	-240
10-20	28	15	15 − 25 = − 10	-280
20-30	32	25 = a	25-25 = 0	0
30-40	25	35	35-25 = 10	250
40-50	13	45	45-25 = 20	260
Total	Σf_i =110			$\Sigma f_i d_i$ = -10

Assumed mean = a = 25

Mean of the data:

$$\bar{x} = a + \frac{\sum f_i d_i}{\sum f_i}$$

= 25 + (-10/ 110)

= 25 -(1/11)

= (275-1)/11

= 274/11

=24.9

Hence, the mean marks of the students are 24.9.

Step-deviation Method

- When the data values are large, the step-deviation method is used to find the mean. The formula is given by:

$$Mean,\ (\overline{x}) = a + h\frac{\sum f_i u_i}{\sum f_i}$$

Here,

a = assumed mean

f_i = frequency of ith class

$x_i - a$ = deviation of ith class

$u_i = (x_i - a)/h$

$\Sigma f_i = N$ = Total number of observations

x_i = class mark = (upper class limit + lower class limit)/2

Example: Using step - deviation method, calculate the mean marks of the following distribution.

Class interval	Frequency
50 – 55	5
55 – 60	20
60 – 65	10
65 – 70	10
70 – 75	9
75 – 80	6
80 – 85	12
85 – 90	8

Solution

Class interval	Mid – value (x_i)	f_i	$d_i = x_i - A$	$u_i = \frac{d_i}{h}$ $(h = 5)$	$f_i \times u_i$
50 – 55	52.5	5	−15	−3	−15
55 – 60	57.5	20	−10	−2	−40
60 – 65	62.5	10	−5	−1	−10
65 – 70	A = 67.5	10	0	0	0
70 – 75	72.5	9	5	1	9
75 – 80	77.5	6	10	2	12
80 – 85	82.5	12	15	3	36
85 – 90	87.5	8	20	4	32

$$\text{Mean } (\bar{x}) = A + \frac{\sum f_i u_i}{\sum f_i} \times h$$

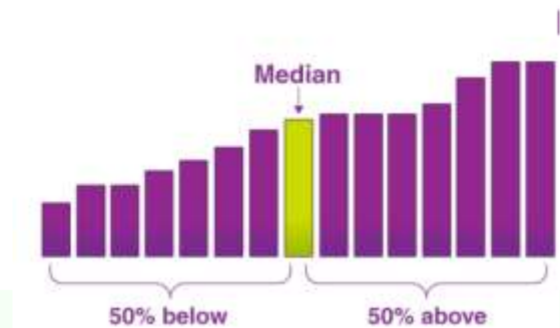
$$= 67.5 + \frac{24}{16}$$

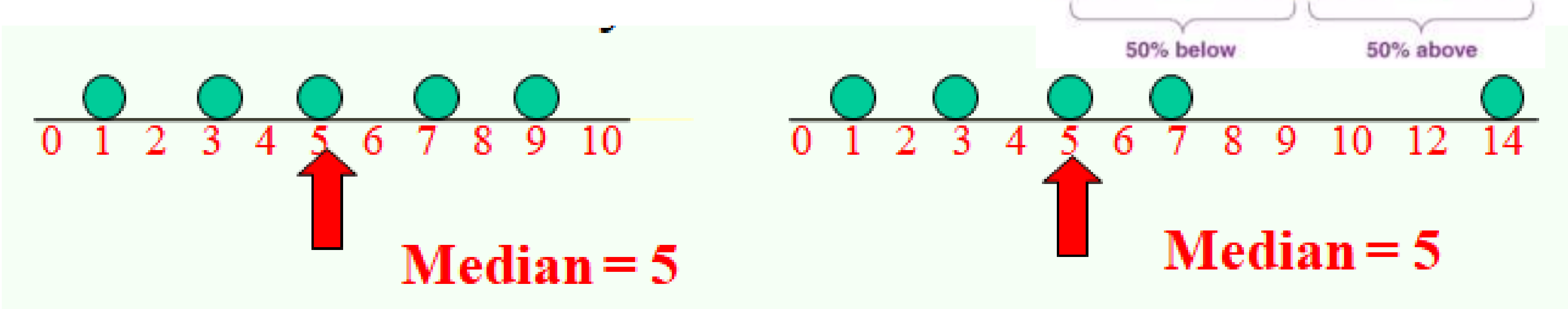
$$= 67.5 + 1.5$$

$$= 69.0$$

Median

- The *median* is simply another name for the 50th percentile
  - It is the score in the middle; half of the scores are larger than the median and half of the scores are smaller than the median
  - Not affected by extreme values





How To Calculate the Median?

The formula to calculate the median of the data set is given as follows.

Odd Number of Observations

If the total number of observations given is odd, then the formula to calculate the median is:

$$Median = (\frac{n+1}{2})^{th}\, term$$

where n is the number of observations

Even Number of Observations

If the total number of observation is even, then the median formula is:

$$Median = \frac{(\frac{n}{2})^{th}\, term + (\frac{n}{2}+1)^{th}\, term}{2}$$

where n is the number of observations

How To Calculate the Median?

To find the median, place all the numbers in ascending order and find the middle.

Example 1:

Find the Median of 14, 63 and 55

solution:

Put them in ascending order: 14, 55, 63

The middle number is 55, so the median is 55.

Example 2:

Determine the median for the given dataset:

4, 7, 3, 8, 6, 2

Solution:

Given dataset: 4, 7, 3, 8, 6, 2

Here, the number of observations is even, i.e., 6 observations are given.

$n = 6$

Now, arrange the numbers in ascending order

2, 3, 4, 6, 7, 8

The formula to calculate the median for odd observations is:

Median = $[(n/2)^{th}$ term + $\{(n/2)+1\}^{th}$ term$]/2$

Median = $[(6/2)^{th}$ term + $\{(6/2)+1\}^{th}$ term$]/2$

Median = $(3^{rd}$ term + 4^{th} term$)/2$

Here, the 3^{rd} term is 4 and the 4^{th} term is 6

Therefore, median = $(4+6)/2$

$= 10/2 = 5$

Therefore, the median for the given dataset is 5.

Calculating the Median for Grouped Data

- In a grouped data, it is not possible to find the median for the given observation by looking at the cumulative frequencies. The middle value of the given data will be in some class interval. So, it is necessary to find the value inside the class interval that divides the whole distribution into two halves. In this scenario, we have to find the 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.

After finding the median class, use the below formula to find the median value.

$$Median = l + \left(\frac{\frac{n}{2} - cf}{f}\right) \times h$$

Where

l is the lower limit of the median class

n is the number of observations

f is the frequency of median class

h is the class size

cf is the cumulative frequency of class preceding the median class.

Now, let us understand how to find the median of a grouped data using the formula with the help of an example.

Example:

The following data represents the survey regarding the heights (in cm) of 51 girls of Class x. Find the median height.

Height (in cm)	Number of Girls
Less than 140	4
Less than 145	11
Less than 150	29
Less than 155	40
Less than 160	46
Less than 165	51

Solution: To find the median height, first, we need to find the class intervals and their corresponding frequencies.

The given distribution is in the form of being less than type,145, 150 …and 165 gives the upper limit. Thus, the class should be below 140, 140-145, 145-150, 150-155, 155-160 and 160-165.

From the given distribution, it is observed that,

4 girls are below 140. Therefore, the frequency of class intervals below 140 is 4.

11 girls are there with heights less than 145, and 4 girls with height less than 140

Hence, the frequency distribution for the class interval 140-145 = 11-4 = 7

Likewise, the frequency of 145 -150= 29 – 11 = 18

Frequency of 150-155 = 40-29 = 11

Frequency of 155 – 160 = 46-40 = 6

Frequency of 160-165 = 51-46 = 5

Therefore, the frequency distribution table along with the cumulative frequencies are given below:

Class Intervals	Frequency	Cumulative Frequency
Below 140	4	4
140 – 145	7	11
145 – 150	18	29
150 – 155	11	40
155 – 160	6	46
160 – 165	5	51

Here, n= 51.

Thus, the observations lie between the class interval 145-150, which is called the median class.

Therefore,

Lower class limit = 145

Class size, h = 5

Frequency of the median class, f = 18

Cumulative frequency of the class preceding the median class, cf = 11.

We know that the formula to find the median of the grouped data is:

$$Median = l + \left(\frac{\frac{n}{2} - cf}{f}\right) \times h$$

Now, substituting the values in the formula, we get

$$Median = 145 + \left(\frac{25.5 - 11}{18}\right) \times 5$$

Median = 145 + (72.5/18)

Median = 145 + 4.03

Median = 149.03.

Therefore, the median height for the given data is 149. 03 cm.

Mode

- A mode is defined as the value that has a higher frequency in a given set of values. It is the value that appears the most number of times.
- **Example**: In the given set of data: 2, 4, 5, 5, 6, 7, the mode of the data set is 5 since it has appeared in the set twice.

Bimodal, Trimodal & Multimodal (More than one mode)

- When there are two modes in a data set, then the set is called **bimodal**
- For example, The mode of Set A = {2,2,2,3,4,4,5,5,5} is 2 and 5, because both 2 and 5 is repeated three times in the given set.
- When there are three modes in a data set, then the set is called **trimodal**
- For example, the mode of set A = {2,2,2,3,4,4,5,5,5,7,8,8,8} is 2, 5 and 8
- When there are four or more modes in a data set, then the set is called **multimodal**

Mode Formula For Grouped Data

- To determine the mode of data in such cases we calculate the modal class. Mode lies inside the modal class. The mode of data is given by the formula:

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

Where,

- l = lower limit of the modal class
- h = size of the class interval
- 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

Example: In a class of 30 students marks obtained by students in mathematics out of 50 is tabulated as below. Calculate the mode of data given.

Marks Obtained	Number of Students
10-20	5
20-30	12
30-40	8
40-50	5

Solution:

The maximum class frequency is 12 and the class interval corresponding to this frequency is 20 – 30. Thus, the modal class is 20 – 30.

Lower limit of the modal class (l) = 20

Size of the class interval (h) = 10

Frequency of the modal class (f_1) = 12

Frequency of the class preceding the modal class (f_0) = 5

Frequency of the class succeeding the modal class (f_2)= 8

$$Mode = l + \left(\frac{f_1 - f_0}{2f_1 - f_0 - f_2}\right) \times h = 20 + \left(\frac{12 - 5}{2 \times 12 - 5 - 8}\right) \times 10 = 26.364$$

RELATION BETWEEN MEAN, MEDIAN AND MODE

- In statistics, for a moderately skewed distribution, there exists a relation between mean, median and mode. This meen-median-mode relationship is known as the **“empirical relationship”** which is defined as **Mode is equal to the difference between 3 times the median and 2 times the mean**. This relation has been discussed in detail below.

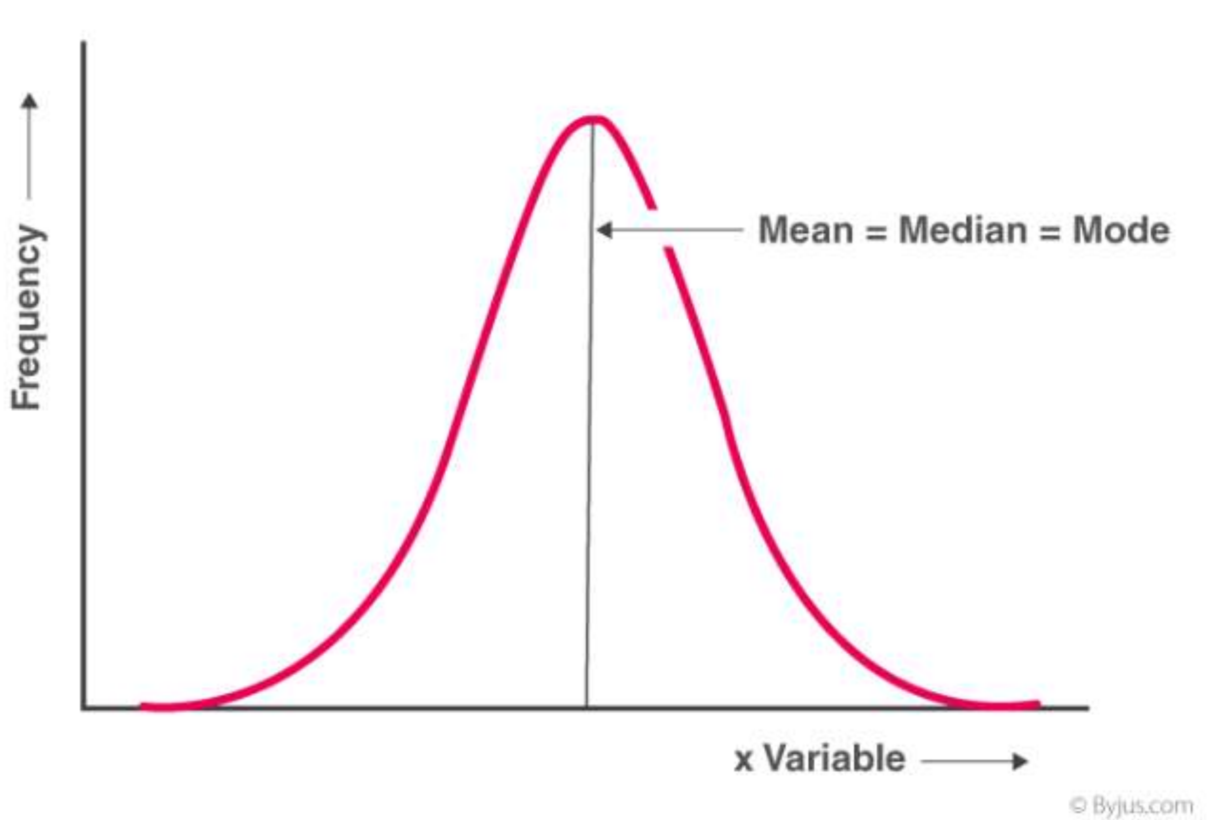
TO RECALL

- **Mean** is the average of the data set which is calculated by adding all the data values together and dividing it by the total number of data sets.
- **Median** is the middle value among the observed set of values and is calculated by arranging the values in ascending order or in descending order and then choosing the middle value.
- **Mode** is the number from a data set which has the highest frequency and is calculated by counting the number of times each data value occurs.

Mean Median Mode Relation With Frequency Distribution

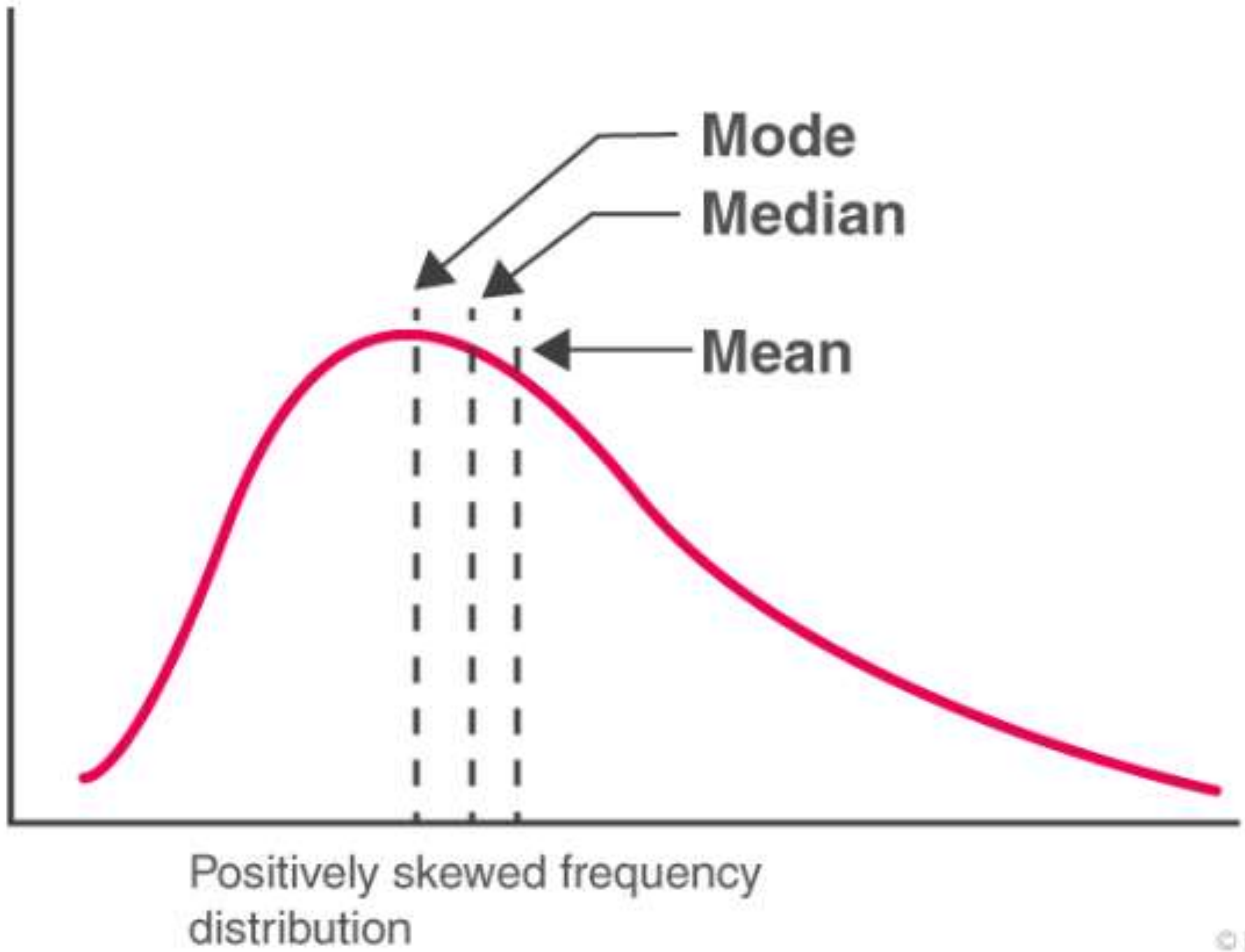
1. Frequency Distribution with Symmetrical Frequency Curve:

Mean = Median = Mode



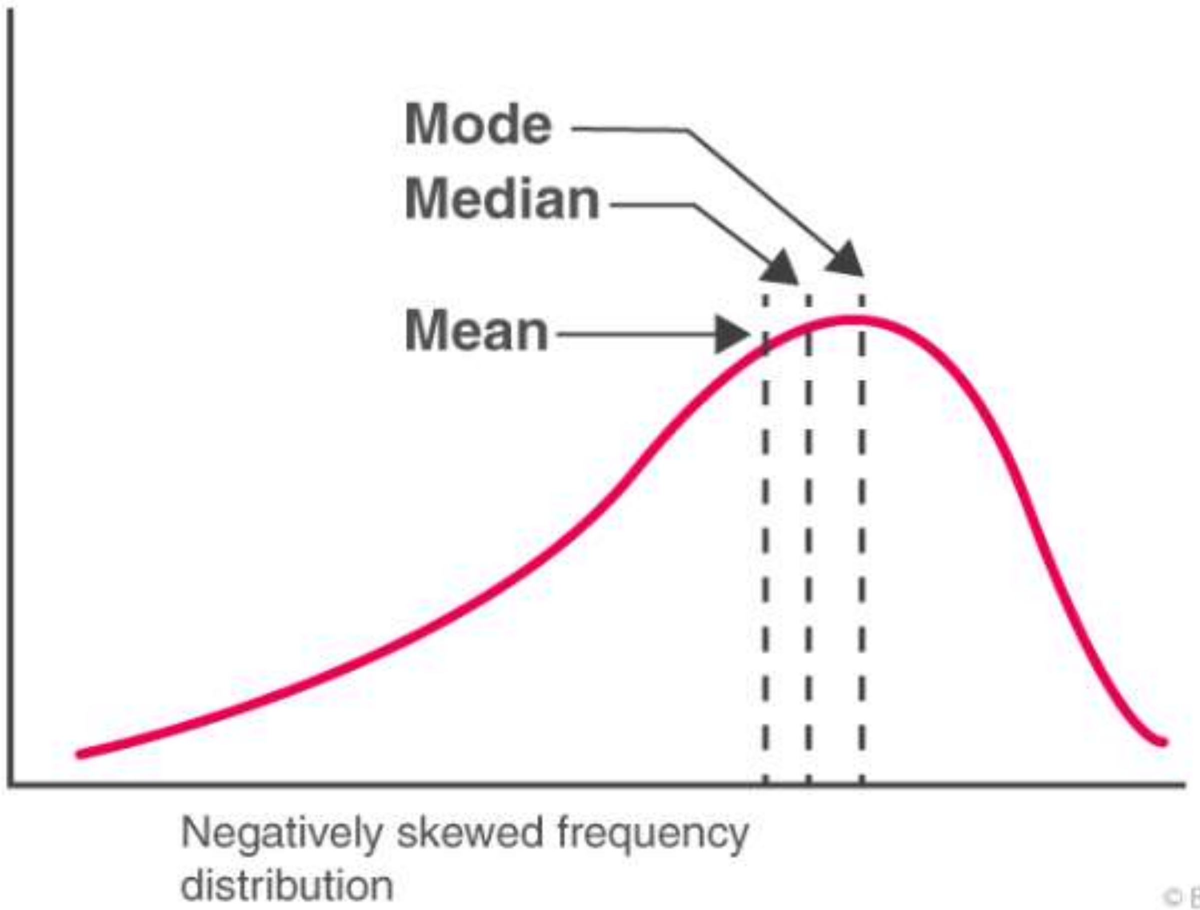
2. For Positively Skewed Frequency Distribution:

Mean > Median > Mode



3. For Negatively Skewed Frequency Distribution

Mean < Median < Mode



EXAMPLE: In a moderately skewed distribution, the median is 20 and the mean is 22.5. Using these values, find the approximate value of the mode.

Solution:

Given,

Mean = 22.5

Median = 20

Mode = x

Now, using the relationship between mean mode and median we get,

(Mean − Mode) = 3 (Mean − Median)

So,

$22.5 - x = 3\,(22.5 - 20)$

$22.5 - x = 7.5$

$\therefore x = 15$

So, Mode = 15.

Mean Median Mode Comparison

<table>
<tr><th>Mean</th><th>Median</th><th>Mode</th></tr>
<tr><td>Mean is the average value that is equal to the ration of sum of values in a data set and total number of values. Mean = Sum of observations/Number of observations</td><td>Median is the central value of given set of values when arranged in an order.</td><td>Mode is the most repetitive value of a given set of values.</td></tr>
<tr><td colspan="3">For example, if we have set of values = 2,2,3,4,5, then;</td></tr>
<tr><td>Mean = (2+2+3+4+5)/5 = 3.2</td><td>Median = 3</td><td>Mode = 2</td></tr>
</table>