

Measures of Central Tendency

Introduction

- One of the most important objectives of statistical analysis is to get one single value that describes the characteristic of the entire mass of data.
- Such a value is called the central value or an average or the expected value of the variable.
- Average is defined as attempt to find a single figure to describe whole of figures.
- The most common measures of central tendency are the *arithmetic mean*, the *median* and the *mode*.

- The goal in measuring central tendency is to describe a group of individuals (more accurately, their scores) with a single measurement.
- Ideally, the value we use to describe the group will be the single value that is most representative of all the individuals.

- Statisticians sometimes use the expression “*number crunching*” to illustrate this aspect of data description.
- That is, we take a distribution consisting of many scores and “*crunch*” them down to a single value that describes them all.

Population and Sample Mean

- Measures of central tendency are defined for a *population data*, denoted by μ (read as “*miu*” and for a *sample data*, denoted by $\bar{x}$ (read as *x bar*).
- **Definition:** It is the most widely used measure for representing the entire data. It is **average** value of all items of a series. It is also called **Mean**. ...

“Arithmetic mean is defined as sum of the different values of an item of a series divided by total number of items.”

Thus Population arithmetic mean is given by:

$$\bar{L} =$$

and the sample arithmetic mean is given by:

$$=$$

Example

- A sample of five executive bankers received the following bonus in 2011 Eth. calendar (in thousands of birr).

47.3, 38.6, 55.0, 25.8, 40.3

Find the average bonus received by the bankers.

$$= \frac{47.3 + 38.6 + 55.0 + 25.8 + 40.3}{5} = 41.4$$

which is 41,400 birr

Types of Arithmetic Mean

Arithmetic mean is of following two types.

1. Simple Arithmetic Mean:

In simple arithmetic mean, all items of a series are given equal importance.

2. Weighted Arithmetic Mean:

Under weighted A.M. different items of the series are assigned weights according to their relative importance.

i.e. weighted mean is the mean of a set of values where each value or measurement has a different weight or degree of importance.

Example: A student scored A in a 3 credit hours course, B in a 5 credit hours course and D in a 4 credit course. This student is going to apply for a competition which require a minimum GPA of 2.5. Check whether a student is going to succeed to fulfil the requirement or not.

The Arithmetic Mean for GFD

The arithmetic mean of a grouped frequency distribution is computed using the following formula:

$\bar{x} = \frac{\sum fx}{N}$, where

- class mark of the class
- frequency of the class

k – the number of classes of the GFD

N – the number of observations in the GFD

Properties of Arithmetic Mean

Suppose we have N observations:

$x_1, x_2, \dots, x_N$ whose mean is given by $\bar{x}$ and let k be any constant number. Then

i) the arithmetic mean of

a) $x_1 + k, x_2 + k, x_3 + k, \dots, x_N + k$ is $\bar{x} + k$

b) $x_1 - k, x_2 - k, x_3 - k, \dots, x_N - k$ is $\bar{x} - k$

c) $\frac{x_1}{k}, \frac{x_2}{k}, \frac{x_3}{k}, \dots, \frac{x_N}{k}$ is $\frac{\bar{x}}{k}$

2. Since $\sum (x_i - \bar{x}) = 0$ it follows that

$$\sum (x_i - \bar{x}) = 0.$$

We use this property in correcting and combining the arithmetic mean.

3. The sum of the deviation of the given observations from the mean is always zero.

i.e. $\sum (x_i - \bar{x}) = 0$.

Correcting the Arithmetic Mean

Example 1:- The arithmetic mean of 75 observations is 46. Later it was found that the two numbers 66 and 85 were misread as 36 and 45. Find the correct arithmetic mean.

Example 2:- The average mark in Calculus I of 70 students is 77. Later it was found that the two marks of two students which are 58 and 81 were left out at the time of calculation. Find the average mark of the 72 students.

Combining two or more averages.

Example 1:- The arithmetic mean of 40 observations is 27 and the arithmetic mean of another 55 observations is 64. Find the arithmetic mean of the 95 observations together.

Example 2:- The average daily wage paid to all male employees in the management group of a company is 255 birr, while the average daily wage of female management group employees is 345 birr. If there are 60 employees in a company's management group and the average daily wage paid to all employees in this management group is 307.50 birr, then find the number of male and female workers of the management group employed in a company.

Merits and demerits AM

Merits:

- Based on all observations
- Easy to calculate and Simple to understand
- Widely used in statistical analysis
- It can be used for comparison and for further mathematical conclusion

Merits and demerits AM (Contd...)

Demerits:

- Can not be computed for qualitative data.
- It is highly affected by extreme values (either low or high).
- Sometimes gives absurd results (4.4 children).
- If a single observation is missing or lost arithmetic mean cannot be calculated.

The Geometric Mean

The geometric mean provides an accurate measure of the average percentage rate of a series of numbers through time.

Definition:- The geometric mean, denoted by GM is found by taking the root of the product of the given *n numbers*.

i.e. $GM =$

Example 1:- Find the AM and GM of 5, 6, 8, 9, 11

Since GM =

=

$$\Rightarrow \log GM = \log$$

$$= \log]$$

$$= \log + \log + \log + \dots + \log$$

$$\Rightarrow GM = \text{Antilog } \log + \log + \log + \dots + \log$$

Example 2:- Consider example 1 and use the logarithm formula to find the GM.

Merits and demerits of GM

Merits:

- It is rigidly defined.
- It is not affected by the extreme items in the series.
- It is based on all the observations of the series.
- It is suitable for measuring the relative changes.
- It is used to find average of the rates of changes.
- It is useful in measuring growth of population
- It is used in averaging the ratios, percentages and in determining the rate gradual increase and decrease.

Merits and demerits of GM (Contd...)

Demerits:

- It is difficult to compute.
- It is not easy to understand.
- If there are negative values or zeros in the series, it can not be computed.

The Harmonic Mean

Definition: - Suppose we have n observations;
, . . .

The Harmonic mean of these n observations, denoted by HM is defined by

$$HM =$$

Example 1:- Find the harmonic mean of 5, 6, 8, 9, 11.

Example 2:- A man travels 300 km each on the three days with speeds of 70 km/hr, 80 km/hr and 90 km/hr. Find the average speed.

Example 3:- A man travels 5 hours each on the three days with speeds of 70 km/hr, 80 km/hr and 90 km/hr. Find the average speed.

Merits and demerits of HM

Merits:

- It is rigidly defined.
- Like AM and GM, it is also based on all the observations of a series i.e. it cannot be calculated ignoring any item of a series.
- It is capable of further algebraic treatment.
- It can be calculated even when a series contains any negative value.
- It is most appropriate average under conditions of wide variations among the items of a series since it gives larger weight to smaller items.
- It is extremely useful while averaging certain types of rates and ratios.

Merits and demerits of HM (Contd...)

Demerits:

- It is difficult to understand and to compute.
- It cannot be computed when one of the values is 0.
- It is necessary to know all the items of a series before it can be calculated.
- It is usually a value which may not be a member of the given set of numbers.

The Median ()

- The median is the middle number in a sorted, ascending or descending, list of numbers and can be more descriptive of that data set than the average.
- The median is sometimes used as opposed to the mean when there are outliers in the sequence that might skew the average of the values.
- If there is an odd amount of numbers, the median value is the number that is in the middle, with the same amount of numbers below and above.
- If there is an even amount of numbers in the list, the middle pair must be determined, added together, and divided by two to find the median value.

Understanding the Median

- Median is the middle number in a sorted list of numbers.
To determine the median value in a sequence of numbers, the numbers must first be sorted, or arranged, in value order from lowest to highest or highest to lowest. The median can be used to determine an approximate average, or [mean](#), but is not to be confused with the actual mean.

Median Example

- To find the median value in a list with an **odd** amount of numbers, one would find the number that is in the middle with an equal amount of numbers on either side of the median. To find the median, first arrange the numbers in order, usually from lowest to highest.
- For example, in a data set of 3, 13, 2, 34, 11, 26, 47, the sorted order becomes 2, 3, 11, 13, 26, 34, 47. The median is the number in the middle 2, 3, 11, **13**, 26, 34, 47, which in this instance is 13 since there are three numbers on either side.

- To find the median value in a list with an **even** amount of numbers, one must determine the middle pair, add them, and divide by two. Again, arrange the numbers in order from lowest to highest.
- For example, in a data set of 3, 13, 2, 34, 11, 17, 27, 47, the sorted order becomes 2, 3, 11, 13, 17, 27, 34, 47. The median is the average of the two numbers in the middle 2, 3, 11, **13**, **17**, 26 34, 47, which in this case is fifteen $(13 + 17) \div 2 = 15$.

Median for ungrouped data

- In the case of ungrouped frequency distribution to compute the median we use the idea of cumulative frequency (of less than type) to arrange the given data in order. Then after we use the formula
$$= \frac{N}{2}, \text{ when } N \text{ is even}$$
and $= \frac{N+1}{2}, \text{ when } N \text{ is odd.}$

Example 1:- The following table refers to the weight (to the nearest Kg) of a sample of students at HiLCoE. Find the median weight and interpret your result.

Weight (in Kg)	Number of students	Cum. Frequency (Less than type)
58	44	
59	136	
60	270	
61	443	
62	211	
63	101	

Median for GFD

In the case of GFD, we first find the median position and identify the class where the median position is found. This class is known as the *median class*. Then to find the value of the median, we use the following formula which we call the *interpolation method*.

$$= LCB + \frac{N}{2} w ,$$

Where *LCB* – Lower class boundary of the
median class.

N – Number of observations

Cumulative frequency of a class just before the
median class.

- The frequency of the median class.

w - The class width of the median class.

Example 1:- The following GFD refers to the weight (to the nearest Kg) of a sample of students at HiLCoE. Find the median weight and interpret your result.

Weight (in Kg)	Number of students	Cum. Frequency (Less than type)
35 - 43	8	
44 - 52	11	
53 - 61	16	
62 - 70	19	
71 - 79	14	
80 - 88	9	

Advantage and disadvantage of median

Advantages of Median:

- It is very simple to understand and easy to calculate. In some cases it is obtained simply by inspection.
- Median lies at the middle part of the series and hence it is not affected by the extreme values.
- In grouped frequency distribution it can be graphically located by drawing ogives.

Advantage and disadvantage of median (Contd ...)

- It is a special average used in qualitative phenomena like intelligence or difficulty or simplicity of an examination which are not quantified but ranks are given. Thus we can locate the person whose intelligence is the average or the weight of an exam giving ranks as very difficult, difficult, fair, simple, very simple and identify the is the average.

Advantage and disadvantage of median (Contd ...)

Disadvantages of Median :

- In simple series, the item values have to be arranged. If the series contains large number of items, then the process becomes tedious.
- It is a less representative average because it does not depend on all the items in the series.
- It is not capable of further algebraic treatment. For example, we can not find a combined median of two or more groups if the median of different groups are given.

Advantage and disadvantage of median (Contd ...)

- It is affected more by sampling fluctuations than the mean as it is concerned with only one item i.e. the middle item.
- It is not rigidly defined. In simple series having even number of items, median cannot be exactly found. Moreover, the interpolation formula applied in the continuous series is based on the unrealistic assumption that the frequency of the median class is evenly spread over the magnitude of the class interval of the median group.

Quartiles, Deciles and Percentiles

By extending the idea of the median, we can think of those values which divide the data set into four equal parts. These values, denoted by Q_1 , Q_2 , and Q_3 are called first (lower), second (middle) and third (upper) quartiles respectively, where Q_2 being equal to the median.



- Similarly the values which divide the data set in to ten equal parts are called deciles and are denoted by
- While the values which divide the data set in to 100 equal parts are called percentiles and are denoted by .
- Again note that the 2nd quartile, the 5th decile and the 50th percentile correspond to the median and the 25th and the 75th percentiles correspond to the 1st and 3rd quartiles respectively.

- To find the positions of quartiles, deciles and percentiles for the given Grouped Frequency Distribution, we use the following formulas and to find the respective values we use the interpolation method as we used in the case of median for the GFD.

=

=

=

Example 1:- The following GFD refers to the weight (to the nearest Kg) of a sample of students at HiLCoE. Find and interpret your result.

Weight (in Kg)	Number of students	Cum. Frequency (Less than type)
35 - 43	8	
44 - 52	11	
53 - 61	16	
62 - 70	19	
71 - 79	14	
80 - 88	9	

The Mode

- The mode is the number that appears most frequently in a set. A set of numbers may have one mode, more than one mode, or no mode at all.
- Uni-modal, Bi-modal and Multi-modal data
- **Example1:-** Find the mode for each of the following observations.

- i) 3, 5, 9, 5, 6, 2, 1 – Uni-modal
- ii) 8, 2, 8, 3, 5, 7, 8, 7 – Uni-modal
- iii) 7, 1, 8, 2, 19 – No mode
- iv) 6, 8, 0, 7, 0, 3, 2 – Uni-modal
- v) 4, 8, 4, 6, 4, 1, 8, 8, 9 – Bi-modal
- vi) 2, 6, 3, 6, 5, 1, 1, 3, 8, 7 – Multi-modal

The Mode for the GFD

Since the definition of mode is the observation that occurs most often, in the case of the GFD, the mode is assumed to be found in that class with the highest frequency which is called the modal class. To find the mode in the case of GFD, we shall use the following formula.

$$= LCB + . W$$

where the frequency of the modal class.

the frequency of the class which is just before
the modal class.

the frequency of the class which is just after
the modal class.

w – the class width of the GFD

$$= LCB + . W$$

where the frequency of the modal class.

the frequency of the class which is just before the modal class.

the frequency of the class which is just after the modal class.

w – the class width of the GFD

Advantages and Disadvantages of the Mode

Advantages:

- The mode is easy to understand and calculate.
- The mode is not affected by extreme values.
- The mode is easy to identify in un-grouped data and discrete [frequency distribution](#).
- The mode is useful for qualitative data.
- The mode can be located graphically.

Advantages and Disadvantages of the Mode (Contd ...)

Disadvantages:

- The mode is not well defined.
- The mode is not based on all values.
- The mode is stable for large values and will not be well defined if the data consist of a small number of values.
- The mode is not capable of further mathematical treatment.
- Sometimes data have one mode, more than one mode, or no mode at all.

Relation Between Mean Median and Mode

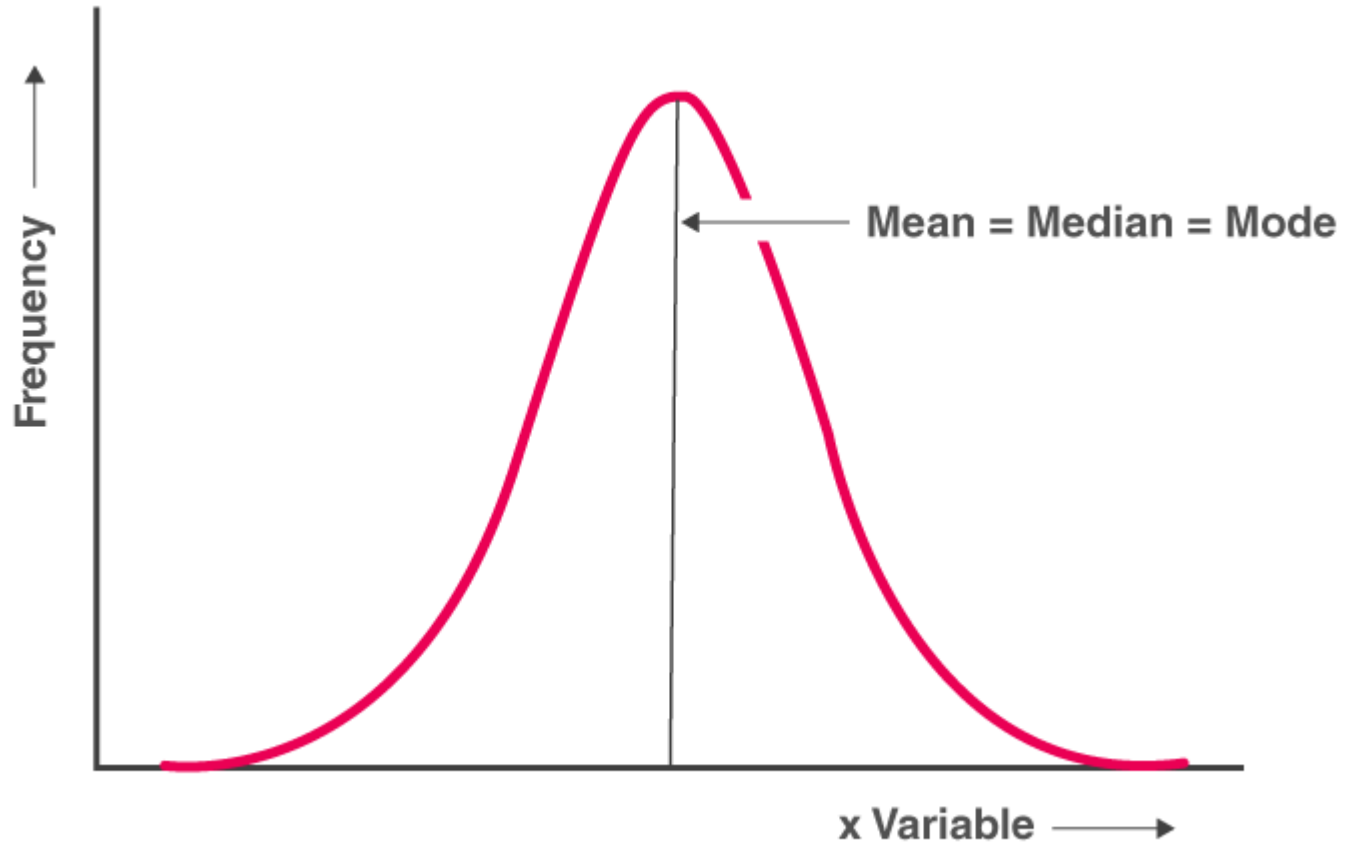
In statistics, for a moderately skewed distribution, there exists a relation between mean, median, and mode. This mean median and mode relationship is known as the “**empirical relationship**” which have been discussed in detail below.

$\frac{2}{3} - \frac{2}{3} [\text{mean} - \text{mode}]$ Empirical Relationship between Mean, Median and Mode

In case of a moderately skewed distribution, the difference between mean and mode is almost equal to three times the difference between the mean and median. Thus, the empirical mean median mode relation is given as:

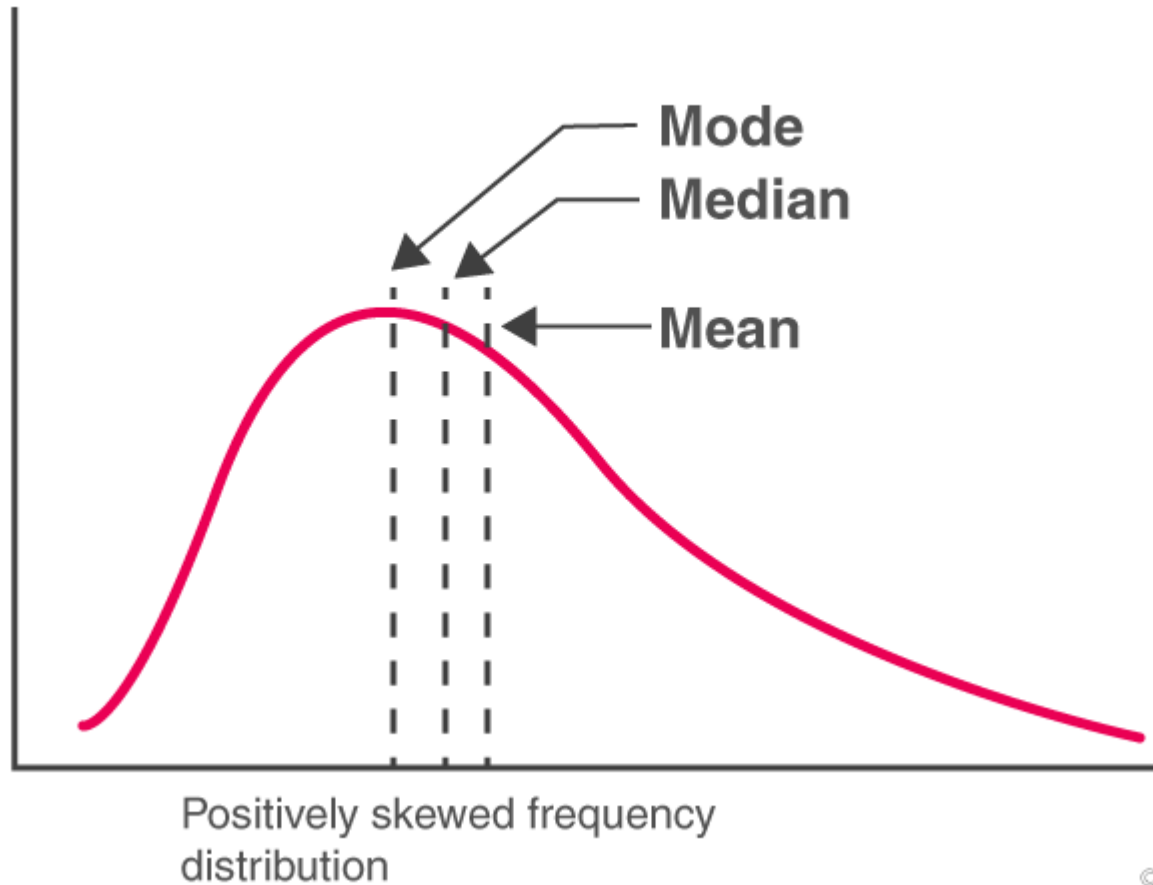
$$\text{Mean} - \text{Mode} = 3 (\text{Mean} - \text{Median})$$

Mean = Median = Mode



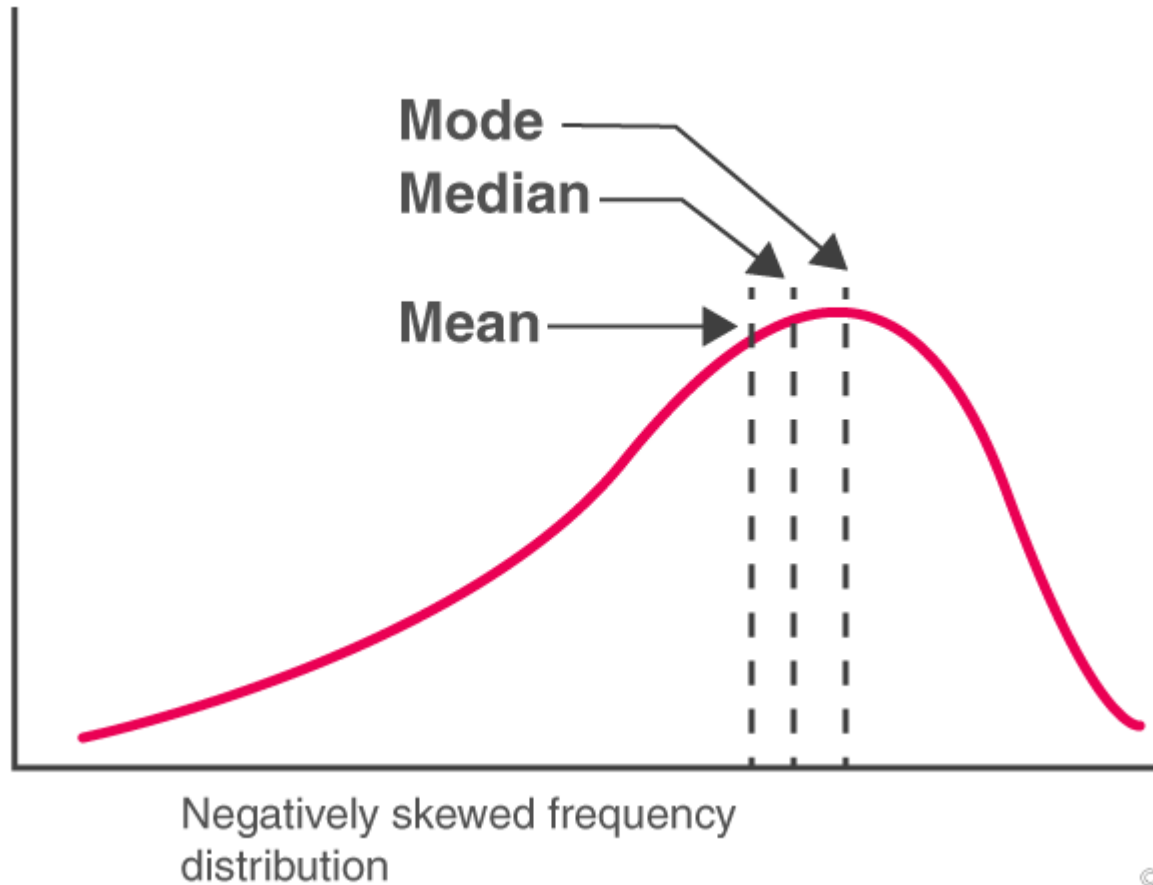
© Byjus.com

Mean > Median > Mode



© Byjus.com

Mean < Median < Mode



© Byjus.com

Question 1: For a moderately skewed distribution, the median is 55 and the mode is 51. Then

(i) the mean is _____

(ii) The type of skewness is_____.

Question2: In a moderately skewed distribution, the median is 20 and the mean is 22.5. Using these values:

(i) find the approximate value of the mode.

(ii) Determine the type of skewness

- **Question3:** In a moderately asymmetrical distribution, the mode is twice the median and the mean is 37. Then the approximate value of the median is _____, the mode is _____ and the type of skewness is _____