# Measures of Central tendency

BS – Unit II

- **Measure of Central Tendency:** It is a single value or figure that represents the entire set of data. It is a value to which most of the observations are closer.
- **Meaning of Arithmetic Mean:**
- Arithmetic Mean is defined as "**the sum of the values of all observations divided by the number of observations"**. It is also known as 'Mean' or 'Average' by the common man. It is generally denoted by $\overline{X}$ if the observations are $\{x_{1,}\ x_2\ \ldots\ldots..\}$

- **Objectives of Averages:**
- (1) To Present a Brief Picture of Data
- Averages summarizes data into a single figure, which makes it easier to understand and remember.
- (2) To Make Comparison Easier
- Averages are very helpful for making comparative studies as they reduce the bulky statistical data to a single figure.
- (3) To Help in Decision-making
- Most of the decisions in research or planning are based on the average value of certain variables.
- (4) To Help in Formulation of Policy
- It is very useful in policy formulation.
- For example: For the removal of poverty from India, government takes into consideration per capita income.

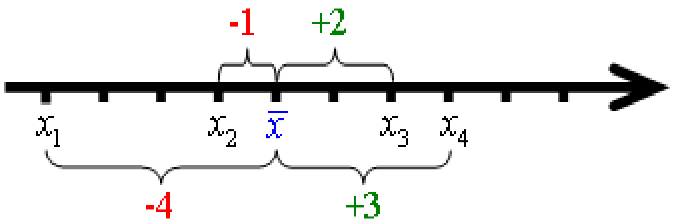
- **Merits and Demerits of Arithmetic Mean**
- **(a) Following Are Some of the Merits of Arithmetic Mean:**
- (1) Easy to Compute
- Its calculation is very easy because it requires knowledge of only simple mathematics i.e. addition, multiplication and division of numbers.
- (2) Simple to Understand
- It is also simple to understand the meaning of arithmetic mean i.e., the value per unit or cost per unit, etc.
- (3) Based on All Items
- It takes into consideration all the values of data.
- It is considered to be more representative of the distribution.

- (4) Rigidly Defined
- Its value is always definite because it is rigidly defined.
- (5) Good Basis of Comparison
- It provides a sound basis of comparison of two or more groups of data.
- (6) Algebraic Treatment
- It is capable of further algebraic treatment. So, it is widely used in advance statistical analysis.

- **Following Are Some of the Demerits of Arithmetic Mean:**
- (1) Complete Data is Required
- It cannot be computed unless all the items of a series are available.
- (2) Affected by Extreme Values
- Since arithmetic average is calculated from all the items of a series, it can be unduly affected by extreme values i.e. very small or very large items.
- (3) Absurd Result
- Sometimes arithmetic mean gives absurd results. For example, if a teacher says that average number of students in a class is 28.75, it sounds illogical.
- (4) Calculation of Mean by Observation Not Possible
- Arithmetic mean cannot be computed by simply observing the series like median or mode.

- (5) No Graphic Representation
- Arithmetic Mean cannot be represented or depicted on graph paper.
- (6) Not Possible in Case of Open Ended Frequency Distribution
- In case of open ended class frequency distribution, it is not possible to compute arithmetic mean without making assumption about the class size.
- Example for open ended classes are <10 or >50 and well defined classes are 0 – 10, 10 – 20 etc.
- (7) Not Possible in Case of Qualitative Characteristics
- It cannot be computed for a qualitative data; like data on intelligence, honesty, smoking habit, etc.

# Arithmetic Mean




In general language arithmetic mean is same as the average of data. It is the representative value of the group of data. Suppose we are given ' n ' number of data and we need to compute the arithmetic mean, all that we need to do is just sum up all the numbers and divide it by the total numbers.

$$\text{Mean} = \frac{Sum of all observations}{number of observations}$$

Thus, the mean of n observation $x_1, x_2, \ldots, x_n$, is given by

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

Where the symbol ∑ called sigma which stands for summation.

# Arithmetic Mean

***(i) Direct Method***

Here we take the mid point of every class interval as X, or $m$ multiply these values of X or $m$ with their respective frequencies. Take total and apply the formula ;

$$\bar{X} = \frac{\Sigma fX}{N} \quad \text{or} \quad \bar{X} = \frac{\Sigma fm}{N}$$

Where $f$ is frequency; X or $m$ the mid point of the class-interval and $N = \Sigma f$

**Example 1. Use direct method to find $\bar{X}$.**

| Income : | 10–20 | 20–30 | 30–40 | 40–50 | 50–60 | 60–70 |
|---|---|---|---|---|---|---|
| No. of Persons : | 4 | 7 | 16 | 20 | 15 | 8 |

**Solution**

| Income | Mid value $m$ | No. of Persons f | fm |
|---|---|---|---|
| 10–20 | 15 | 4 | 60 |
| 20–30 | 25 | 7 | 175 |
| 30–40 | 35 | 16 | 560 |
| 40–50 | 45 | 20 | 900 |
| 50–60 | 55 | 15 | 825 |
| 60–70 | 65 | 8 | 520 |
| | | N = 70 | $\Sigma fm = 3040$ |

As $\bar{X} = \frac{\Sigma fm}{N}$

$\therefore \quad \bar{X} = \frac{3040}{70} = 43.43.$

### (ii) *Short Cut Method*

Here also the mid point of each interval is taken as X or $m$ and then its deviations are taken from Assumed mean. Then the following formula is applied.

$$\overline{X} = A + \frac{\Sigma fdx}{N}$$

Where A is Assumed Mean ; $f$ the frequency $dx$, the deviation from mean and $N = \Sigma f$.

**Steps to Calculate**

(*a*) Find mid points of each class as X or $m$.

(*b*) Take an assumed mean A.

(*c*) From the mid-point of each class, deduct the assumed mean $(m-A)$ and find out deviations $dx$.

(*d*) Multiply $f$ by respective $dx$ to find $fdx$ and eventually $\Sigma fdx$.

(*e*) Apply the above formula

**Example 2.** Use short cut method to find $\overline{X}$ for the data given in example 1.

## Solution

Let Assumed Mean (A) be = 45.

| Income | Mid Value $m$ | No. of Persons f | $(m-A) = dx$ | fdx. | |
|---|---|---|---|---|---|
| 10–20 | 15 | 4 | –30 | –120 | |
| 20–30 | 25 | 7 | –20 | –140 | –420 |
| 30–40 | 35 | 16 | –10 | –160 | |
| 40–50 | 45 | 20 | 0 | 0 | |
| 50–60 | 55 | 15 | 10 | 150 | +310 |
| 60–70 | 65 | 8 | 20 | 160 | |
| | | N = 70 | | $\Sigma fdx = -110$ | |

As $\overline{X} = A + \frac{\Sigma fdx}{N}$

$\therefore \quad \overline{X} = 45 + \frac{-110}{70}$

$= 45 - 1.57 = \mathbf{43.43}$

(**Note** :–A is taken from among mid values $m$. If number of mid values or class intervals is odd, then central value is taken as mid-point ; And if is even, we take any of the two middle values as this value.

## (iii) *Step-deviation Method*

It is the most simplified method, with short calculations. A common factor is taken and all the deviations *dx* are divided by that factor, say *i*; and then the following formula is applied.

$$\overline{X} = A + \frac{\Sigma fd'x}{N} \times i$$

Where A is Assumed mean ; *f* the frequency :

N= $\Sigma f$, *i* is common factor such that $d'x = \frac{dx}{i}$ ...(*i*)

**Example 3. Use step deviation method to find $\overline{X}$ for the data given in example 1.**

**Solution** Let Assumed Mean (A) be = 45

| Income X | Mid value m | f | $d'x = \frac{m - A}{i}$ | fdx′ |
|---|---|---|---|---|
| 10–20 | 15 | 4 | $\frac{15-45}{10} = -3$ | −12 |
| 20–30 | 25 | 7 | $\frac{25-45}{10} = -2$ | −14 −42 |
| 30–40 | 35 | 16 | $\frac{35-45}{10} = -1$ | −16 |
| 40–50 | 45 | 20 | $\frac{45-45}{10} = 0$ | 0 |
| 50–60 | 55 | 15 | $\frac{55-45}{10} = 1$ | 15 31 |
| 60–70 | 65 | 8 | $\frac{65-45}{10} = 2$ | 16 |
| | | N = 70 | | $\Sigma fd'x = -11$ |

$$\overline{X} = A + \frac{\Sigma fd'x}{N} \times i$$

$$\therefore \overline{X} = 45 + \frac{-11}{70} \times 10$$

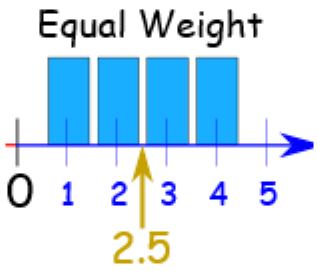
$$= 45 - \frac{11}{7} = 45\text{–}1.57 = \mathbf{43.43.}$$

**Note** : If length of class intervals is equal then we can take *d′x* direclyas in this problem or in the next problems.

**Mean**

When we do a simple mean (or average), we give equal weight to each number.

Here is the mean of 1, 2, 3 and 4:



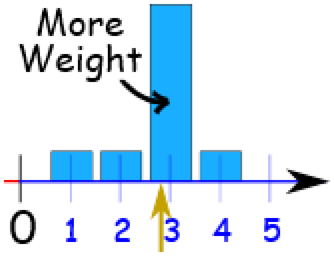

Add up the numbers, divide by how many numbers:

Mean = *(1 + 2 + 3 + 4)* / **4** = *10* / **4** = 2.5

- **Weights**
- We could think that each of those numbers has a "weight" of ¼ (because there are 4 numbers)
- Mean $= \frac{1}{4} \times 1 + \frac{1}{4} \times 2 + \frac{1}{4} \times 3 + \frac{1}{4} \times 4$

  $= 0.25 + 0.5 + 0.75 + 1 = \mathbf{2.5}$
- Same answer. As long as the weights are same for all numbers mean is same.

Now let's change the weight of **3** to 0.7, and the weights of the other numbers to 0.1 so **the total of the weights is still 1**:




Mean = 0.1 × 1 + 0.1 × 2 + 0.7 × 3 + 0.1 × 4

= 0.1 + 0.2 + 2.1 + 0.4 = **2.8**

This **weighted mean** is now a little higher ("pulled" there by the weight of 3).

**Decisions**

Weighted means can help with decisions where some things are more important than others:

Example: Sam wants to buy a new camera, and decides on the following rating system:

- Image Quality **50%**
- Battery Life **30%**
- Zoom Range **20%**

The Sonu camera gets 8 (out of 10) for Image Quality, 6 for Battery Life and 7 for Zoom Range

The Conan camera gets 9 for Image Quality, 4 for Battery Life and 6 for Zoom Range

Which camera is best?

$$\text{Sonu: } 0.5 \times 8 + 0.3 \times 6 + 0.2 \times 7 = 4 + 1.8 + 1.4 = \mathbf{7.2}$$

$$\text{Conan: } 0.5 \times 9 + 0.3 \times 4 + 0.2 \times 6 = 4.5 + 1.2 + 1.2 = \mathbf{6.9}$$

Sam decides to buy the Sonu.

## Summary

- **Weighted Mean**: A mean where some values contribute more than others.
- When the weights add to 1: just multiply each weight by the matching value and sum it all up
- Otherwise, multiply each weight **w** by its matching value **x**, sum that all up, and divide by the sum of weights:
- Weighted Mean = $\Sigma w\, x \,/\, \Sigma \mathbf{w}$

# Combined mean

A combined mean is **a mean of two or more separate groups**, and is found by :
Calculating the mean of each group,
Combining the results.
**Combined Mean Formula**
More formally, a combined mean for two sets can be calculated by the formula

$$x_c = \frac{m \cdot x_a + n \cdot x_b}{m + n}$$

Where:
$x_a$ = the mean of the first set,
m = the number of items in the first set,
$x_b$ = the mean of the second set,
n = the number of items in the second set,
$x_c$ the combined mean.

- Suppose you are running a survey on math proficiency (as measured by an achievement test) in kindergarten, and you have results from two different schools.
- In school 1, 57 kindergarteners were tested and their mean score was 82.
- In school 2, 23 kindergartners were tested and their mean score was 63.
- The combined mean can be calculated by plugging in our numbers into the formula given above:
- [(57*82)+(23*63)]/(57+23) = 76.5.

Now suppose you were running a survey on reading speed, as measured by how long it took 1st graders to read a given block of text. Your results come in for five schools:

| School ID | Number of Students | Average Time |
|---|---|---|
| School 1 | 189 | 83 |
| School 2 | 46 | 121 |
| School 3 | 89 | 82 |
| School 4 | 50 | 147 |
| School 5 | 12 | 60 |

To calculate the combined mean:

1. Multiply column 2 and column 3 for each row,
2. Add up the results from Step 1,
3. Divide the sum from Step 2 by the sum of column 2.

(189*83+46*121 +89*82 +40*147+12*60)/(189+46+89+50+12)

Plug that in your calculator, and the answer you get—91.06—is the combined mean for all five schools; the average reading time for all students.

This same method may be used to combine any number of means.

# Geometric Mean

**Geometric mean** involves roots and multiplication, not addition and division. You get geometric mean by multiplying numbers together and then finding the **n**th root of the numbers such that the nth root is equal to the amount of numbers you multiplied. Geometric mean is useful in many circumstances, especially problems involving money.

For example, if you multiply *three* numbers, the geometric mean is the *third* root of the product of those three numbers. The geometric mean of *five* numbers is the *fifth* root of their product.

- **Geometric Mean of Ungrouped Data**
- If the data is not presented in the frequency distribution then the geometric mean can be calculated by simply taking the logarithm of all observations, adding them up, dividing them by total number of observations and taking antilog of the resultant number. The calculation of geometric mean can be elaborated with the help of following problem.
- **Problem:** A student got the following scores (out of 100) in different subjects in an examination: Urdu: 8, English: 67, Biology: 85, Chemistry: 35, Physics: 50, Math: 100. Find the geometric mean.

| $x$ | $\log x$ |
|---|---|
| 8 | 0.90309 |
| 67 | 1.82607 |
| 85 | 1.92941 |
| 35 | 1.54406 |
| 50 | 1.69897 |
| 100 | 2.00000 |
| | 9.901622 |

$$G.M = Antilog\left(\frac{\sum \log x}{N}\right)$$

$$G.M = Antilog\left(\frac{9.901622}{6}\right)$$

$$G.M = Antilog\,(1.65026) = 44.69$$

- **Geometric Mean of Grouped Data**
- If the data is presented in the frequency distribution then we have to first calculate the mid points of each class interval, take antilog of the mid points, multiply them with the corresponding frequencies, add the resulting values, divide them by sum of frequencies and take antilog of the final number. Geometric mean calculation of grouped data can be elaborated with the help of following problem.

# G. M (Grouped distribution)

| Class Interval | 5 -- 7 | 8 -- 10 | 11 -- 13 | 14 -- 16 | 17 -- 19 |
|---|---|---|---|---|---|
| Frequencies | 15 | 18 | 27 | 10 | 6 |

| *Class Interval* | $f$ | $x$ | $\log x$ | $f \log x$ |
|---|---|---|---|---|
| 5 -- 7 | 15 | 6 | 0.77815125 | 11.67226876 |
| 8 -- 10 | 18 | 9 | 0.95424251 | 17.17636517 |
| 11 -- 13 | 27 | 12 | 1.07918125 | 29.13789364 |
| 14 -- 16 | 10 | 15 | 1.17609126 | 11.76091259 |
| 17 -- 19 | 6 | 18 | 1.25527251 | 7.531635031 |
| | 76 | | | 77.2791 |

$x = Midpoint$

$$x = \frac{L.C.L + U.C.L}{2}$$

$$e.g\ x = \frac{5+7}{2} = 6$$

$$G.M = Antilog\left(\frac{\sum f \log x}{N}\right)$$

$$G.M = Antilog\left(\frac{77.2791}{76}\right)$$

$$G.M = Antilog\,(1.0168299) = 10.40$$

# Harmonic Mean

The harmonic mean of a set of observations is the **reciprocal** of the **arithmetic mean** of the **reciprocal** of the observations. Harmonic mean is defined only for non-zero positive values and is used for averaging while keeping one variable constant. For example in first test a typist types 400 words in 50 minutes, in second test he types the same words (400) in 40 minutes and in third test he takes 30 minutes to type the 400 words. Determine his average speed for the three tests. (Note here constant variable is 400 words).

Therefore,

$$H.M = \frac{N}{\sum 1/x} = \frac{3}{\frac{1}{50}+\frac{1}{40}+\frac{1}{30}} = \frac{3}{0.078} = 38.3\ minutes.$$

The calculation of H.M for the ungrouped data is slightly different from that of grouped data. Calculations for both types of data are given below.

**Harmonic Mean of Ungrouped Data**

In order to calculate the harmonic mean of ungrouped data the formula will be:

$$H.M = \frac{N}{\Sigma 1/x}$$

Here,

N is the total number of observations

And x represents the values of observations while Σ (1/x )shows the sum of reciprocal of observations.

# H.M (Ungrouped data)

**Problem:** During one month the total number of kilometers driven by each truck is given below. Find the geometric mean.

| Truck Number | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| Km. driven | 40 | 50 | 60 | 75 |

| $x$ | $1/x$ |
|---|---|
| 40 | 0.02500 |
| 50 | 0.02000 |
| 60 | 0.01677 |
| 75 | 0.01333 |
| | 0.07500 |

$$H.M = \frac{N}{\sum 1/x}$$

$$H.M = \frac{4}{0.07500}$$

$$H.M = 53.33\,Km.$$

# Harmonic Mean of grouped Data

Harmonic mean of grouped data can be calculated with the help of following formula:

$$H.M = \frac{N}{\Sigma f\,/x}$$

Here,
N is the sum of all frequencies
f is the frequency corresponding to each observation ‘ x’ while Σ(f/x) represents the sum of reciprocal of grouped observations.

# H.M - Example

| Class Interval | 11 -- 15 | 16 -- 20 | 21 -- 25 | 26 -- 30 | 31 -- 35 |
|---|---|---|---|---|---|
| Frequencies | 15 | 20 | 60 | 150 | 15 |

| *Class Interval* | *x* | *f* | *f/x* |
|---|---|---|---|
| 11 – 15 | 13 | 15 | 1.153846154 |
| 16 – 20 | 18 | 20 | 1.111111111 |
| 21 – 25 | 23 | 60 | 2.608695652 |
| 26 – 30 | 28 | 150 | 5.357142857 |
| 31 – 35 | 33 | 15 | 0.454545455 |
| | | 260 | 10.68534 |

$x = Midpoint$

$$x = \frac{L.C.L + U.C.L}{2}$$

$$e.g\ x = \frac{11 + 15}{2} = 13$$

$$H.M = \frac{N}{\sum f/x}$$

$$H.M = \frac{260}{10.68534} = 24.3$$

# Median

- A median is a positional number that determines the position of the middle value of data. It divides the set of data into two equal parts. In which, one part includes all the greater values or which is equal to a median value and the other set includes all lesser values or equal to the median. In simple words, the median is the middle value when a data set is organized according to the magnitude. The value of the median remains unchanged if the size of the largest value increases because it is defined by the position of various values.
- To evaluate the median, the observations must be arranged in the increasing order or decreasing order (data array). For instance, while evaluating the median if there are odd number of observations , the median will be the middle value, with equal number of observations presented below or above. However, if the number of observations are even then median is the average of middle two observations i.e. the middle pair must be evaluated, combined together, and divided by two to find the median value.

- **Meaning, Merits and Demerits of Median**
- **MEDIAN** “The median is that value of the variable which divides the group into two equal parts, one part comprising all values greater and the other values less than the median.”….L.R. Connor
- Median is the middle value of the series when items are arranged either in ascending or descending order.
- It divides the series into two equal parts. One part comprises all values greater than the median and the other part comprises all values smaller than the median.

- Merits of Median
- **(1) EASY TO CALCULATE AND SIMPLE TO UNDERSTAND:**
- It is easy to calculate and simple to understand.
- In many situations median can be located simply by inspection.

- **(2) NOT AFFECTED BY EXTREME VALUES**
- It is not affected by the extreme values i.e. the largest and smallest values. Because it is a positional average and not dependent on magnitude.

- **(3) RIGIDLY DEFINED**
- It has a definite and certain value because it is rigidly defined.

- **(4) BEST AVERAGE IN CASE OF QUALITATIVE DATA**
- Median is the best measure of central tendency when we deal with qualitative data, where ranking is preferred instead of measurement or counting.

- **(5) USEFUL IN CASE OF OPEN ENDED DISTRIBUTION**
- It can be calculated even if the value of the extremes is not known. But the number of items should be known.

- **(6) REPRESENTED GRAPHICALLY**
- Its value can be determined or represented graphically with the help of Ogive curves. Whereas it is not possible in case of Arithmetic Mean.

- **Demerits of Median**
- **(1) ARRANGEMENT OF DATA IS NECESSARY**
- Since the median is an average of position, therefore arranging the data in ascending or descending order of magnitude is time-consuming in the case of a large number of observations.
- **(2) NOT BASED ON ALL THE OBSERVATIONS**
- It is a positional average and doesn’t consider the magnitude of the items.
- It neglects the extreme values.
- **(3) NOT A REPRESENTIVE OF THE UNIVERSE**
- It is not dependent on all the observations so, it cannot be considered as good representative of data.
- In case there is a big variation between the data, it will not be able to represent the data.

- **(4) AFFECTED BY FLUCTUATIONS IN SAMPLING**
- It is affected by the fluctuations of sampling and this effect is more than in case of Arithmetic Mean.

- **(5) LACK OF FURTHER ALGEBRAIC TREATMENT**
- It is a positional average so further algebraic treatment is not possible. Like, we cannot compute the combined median of two groups of data.

**Median of Grouped Data**

Median is a measure of central tendency which gives the value of the middle-most observation in the data. In case of ungrouped data, we first arrange the data values of the observations in ascending order. Then, if n is odd, the median is the (n+1)/2 th observation. But in case of Grouped data, it is difficult to find (n+1)/2 th observation. We use formula to find Median.

We first find cumulative frequency & then locate the class whose cumulative frequency is greater than (and nearest to) n/2 , where n is total observations. This is called the median class

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

- where l = lower limit of median class,
- n = number of observations,
- cf = cumulative frequency of class preceding the median class,
- f = frequency of median class,
- h = class size / width (assuming class size to be equal).

**Numerical**: The following frequency distribution gives the monthly consumption of electricity of 68 consumers of a locality. Find the median.

| Monthly Consum. | 65-85 | 85-105 | 105-125 | 125-145 | 145-165 | 165-185 | 185-205 |
|---|---|---|---|---|---|---|---|
| No. of consumers | 4 | 5 | 13 | 20 | 14 | 8 | 4 |

Solution:

Step 1: Find Cumulative Frequency

Step 2: Find Median Class

Step 3: Apply formula to find Median

h=145 – 125 = 20

| Consumer | freq | Cum freq |
|---|---|---|
| 65-85 | 4 | 4 |
| 85-105 | 5 | 9 |
| 105-125 | 13 | 22 |
| 125-145 | 20 | 42 |
| 145-165 | 14 | 56 |
| 165-185 | 8 | 64 |
| 185-205 | 4 | 68 |

l=125

f=20

cf=22

Median Class

n=68

Applying the formula, Median = 125 + (34 -22)/20 * 20 = 137

- **PARTITIONAL VALUES** :Partition values are the values which are obtained by dividing a series into more than two parts.
- **Quartiles**
- Quartile divides a series into four equal parts.
- For any series, there will be three quartiles.
- **First Quartile** also known as **Lower Quartile** ($Q_1$):
- It divides the distribution in such a way that the one-fourth (25%) of total items fall below it and three-fourth (75%) are above it.
- **Second Quartile** ($Q_2$) or Median: It divides the distribution in two equal halves.
- **Third Quartile** also known as **Upper Quartile** ($Q_3$):
- It divides the distribution in such a way that three-fourth (75%) of total items fall below it and one-fourth (25%) are above it.

# Assignment

- **Q.1 The ________ is that value of the variable which divides the group into two equal parts.**
- a. Mean
  b. Mode
  c. Median
  d. Both (a) and (c)
- **Q.2 Which of the following is merits of Median value of data?**
- a. Easy To Calculate And Simple To Understand
  b. Rigidly Defined
  c. Not affected by extreme values
  d. All of the above
- **Q.3 Which of the following is demerits of Median value of data?**
- a. Arrangement of data is mandatory
  b. Affected by fluctuation in Sampling
  c. Lack of further algebraic treatment
  d. All of the above

- **Q.4 Median _______ extreme values.**
- a. includes
  b. does not includes
  c. rejects
  d. None of the above
- **Q.5 Median is not dependent upon which of the following criteria?**
- a. All observations
  b. Extreme values
  c. Least values
  d. All of the above
- **Answer Key** 1-a, 2-d, 3-d, 4-b, 5-d

# Mode

- The measure of central tendency mode is the value that appears regularly in the data set. On a histogram or a bar chart, the highest bar in the chart is the mode. In the data set if the data has multiple values and has occurred multiple times then the data has a mode. If the data have no value repeats than it does not have a mode.
- Typically, the mode is used with ordinal, category, and discrete data. Also, the mode is only the measure that uses category data- for instance, the most liked flavored ice-cream. But, the category data doesn’t have a central value because it is not possible to order the group. However, the ordinal and discrete data has a mode with value and which is not in the center. In simple words, mode represents the most common value.

- **MODE**
- Mode is that value which occurs most frequently in a distribution.
- It is the most common value found in a series.
- It is that value of the variable which has the highest frequency.

- **Mode**
- Mode is the value which occurs most frequently in a set of observations. For example, {6, 3, 9, 6, 6, 5, 9, 3} the Mode is 6, as it occurs most often.
- **Properties of Mode :**

1. Sometimes there can be more than one mode. Having two modes is called *bi - modal*. Having more than two modes is called *multimodal*.
2. There is an empirical relationship between Mean, Median, and Mode.

- **Mean – Mode = 3 [ Mean – Median ]**

3. Mode can be useful for qualitative data.

4. Mode can be located graphically.

5. Mode can be computed in an open-end frequency table

.

6. Mode is not affected by extremely large or small values.

**Formula for Mode of grouped data :**

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

Where

$l$ ⟶ Lower Boundary of modal class

$h$ ⟶ Size of model class

$f_m$ ⟶ Frequency corresponding to modal class

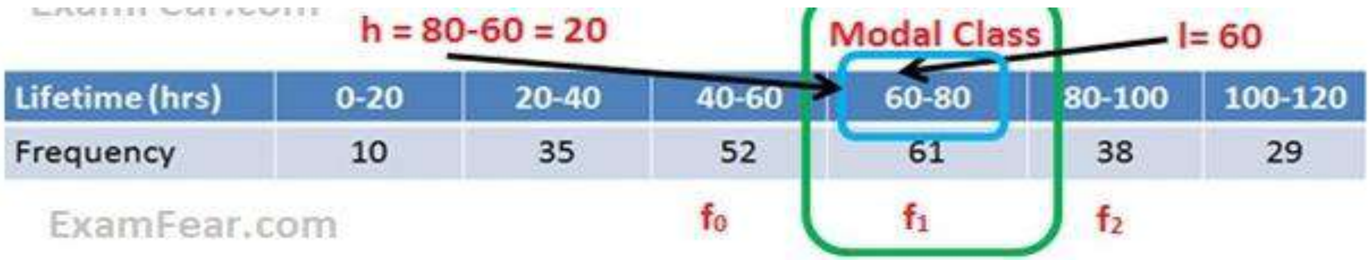
$f_1$ ⟶ Frequency preceding to modal class

$f_2$ ⟶ Frequency proceeding to modal class

E.g. Looking at data below, we can say that maximum occurrence occur at class 60-80, frequency 61. But we can't tell the most frequent data (mode).

| Lifetime (hrs) | 0-20 | 20-40 | 40-60 | 60-80 | 80-100 | 100-120 |
|---|---|---|---|---|---|---|
| Frequency | 10 | 35 | 52 | 61 | 38 | 29 |

We can use this formula to find the mode for Grouped data. In the example above modal class is 60-80.



| Lifetime (hrs) | 0-20 | 20-40 | 40-60 | 60-80 | 80-100 | 100-120 |
|---|---|---|---|---|---|---|
| Frequency | 10 | 35 | 52 | 61 | 38 | 29 |




For the example above, modal class is 60-80 with frequency 61

Therefore, l=60, h = 80-60 = 20, $f_1$ =61, $f_0$ = 52, $f_2$ = 38

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



Mode = 60 + (61-52)/(2*61-52-38) *20

Or Mode = 65.625

- Merits of Mode
- **(1) EASY TO CALCULATE & SIMPLE TO UNDERSTAND**
- It is very easy to calculate.
- In some cases it can be determined just by observation or inspection.
- Everyone understands the concept of majority. Mode is based on this concept so, it’s easy to understand.
- **(2)REPRESENTATIVE VALUE**
- It is a value around which there is maximum concentration of observations.
- Hence, it can be considered as the best representative of the data.

- **(3) NOT AFFECTED BY THE VALUE OF EXTREME ITEMS** It is not affected by extreme values of the given data.
- It can be calculated even if these extreme observations are not known.
- **(4) NO NEED OF COMPLETE DATA** We can find mode even in case of open ended frequency distribution.
- We basically need the point of maximum concentration of frequencies, it is not necessary to know all the values.

- **(5) USEFUL FOR BOTH QUANTITATIVE & QUALITATIVE DATA**
- It can be used to describe quantitative as well as qualitative data.
- For example: In the surveys it is used to measure taste and preferences of people for a particular brand of the commodity.
- **(6) GRAPHIC DETERMINATION**
- It can be determined graphically with the help of Histogram.

# Demerits of Mode

- **(1) NOT BASED ON ALL THE OBSERVATIONS OF THE SERIES**
- The value of mode is not based on each and every item of the series as it considers only the highest concentration of frequencies.
- **(2) SOMETIMES IT IS INDETERMINATE OR ILL DEFINED**
- Value of mode may not be determined always.
- Some distributions can be Bi-modal, Tri-modal or Multi-modal.

- **(3) NOT RIGIDLY DEFINED**
- There are two methods of determining mode, Inspection Method and Grouping Method. We may not get the same value of mode by the two methods. So, it is not rigidly defined.
- **(4) AFFECTED BY THE FLUCTUATIONS OF SAMPLING**
- Mode is affected by sampling fluctuations to a great extent.
- This effect is more than that in case of Mean.

- **(5) COMPLEX GROUPING PROCESS**
- Grouping of data is desirable for correct computation but it is a complex process and involves so much calculations.

- **(6) NOT CAPABLE OF ALGEBRAIC TREATMENT**
- Since it is not based on all the observations and not rigidly defined, it is not suitable for further algebraic treatment.

# Assignment

- **Q.1 __________ is that value which occurs most frequently in a distribution.**
- a. Mean
- b. Median
- c. Mode
- d. None of the above

- **Q.2 Which of the following is a merit of calculating mode?**
- a. Easy to calculate
- b. Simple to understand
- c. No need of complete data
- d. All of the above

- **Q.3 Which of the following is a demerit of calculating mode?**
- a. Not based on all observation
- b. Not rigidly defined
- c. Complex grouping process
- d. All of the above
- **Answer Key 1-a, 2-d, 3-d**