Measures Of Central Tendency

central value

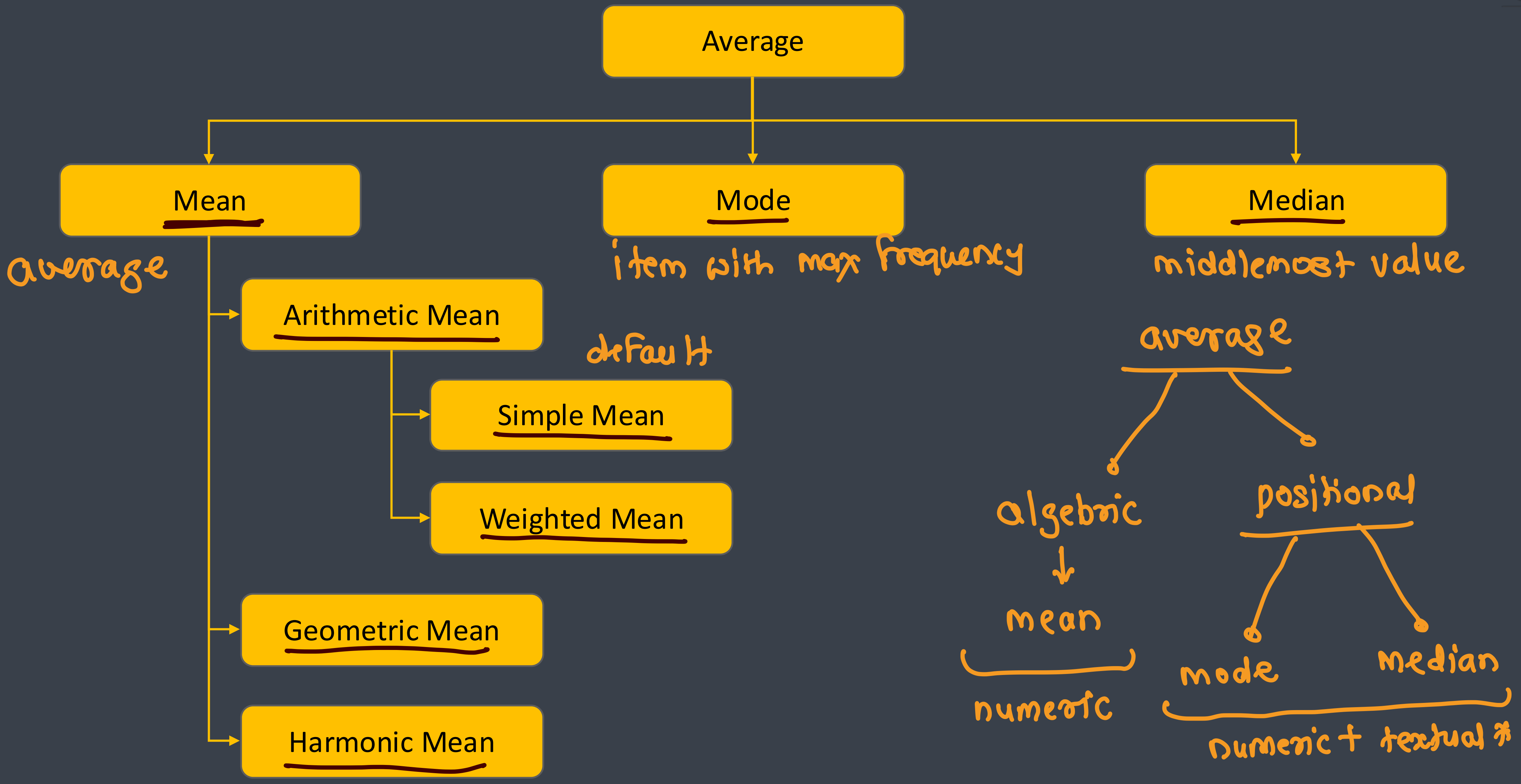
Measures of Central Tendency

- One of the important objectives of statistical analysis is to get one single value that describes the characteristic of entire mass of selected data
- Such value is called as "Central Value" or "Average" or expected value of the variable
- Average
  - Average is an attempt to find one single figure to describe the whole of figures
  - Average is a single value selected from a group of values to represent them in some way
  - Average is sometimes described as a number which is typical of the whole group
- Objectives of averaging
  - To get single value that describes the characteristics of the entire group
  - To facilitate comparison

Requisites of good average

- Easy to understand
- Simple to compute
- Based on all the items **
- Not be unduly affected by extreme observations → outliers
- Rigidly defined → formula
- Capable of further algebraic treatment
- Sampling stability

Types of Averages



Mean

Simple Arithmetic Mean – Individual Series

- Direct method
- Steps
  - Add all the observations together and obtain the total $\sum X$
  - Divide the total by number of observations

$$\bar{X} = \frac{X1 + X2 + X3\ \ldots + Xn}{N}$$

OR

$$\bar{X} = \frac{\sum X}{N}$$

Simple Arithmetic Mean – Individual Series

- Shortcut method (Using Assumed Mean)
- Steps
  - Take an assumed mean and denote it as A
  - Take the deviations of items from assumed mean and denote them by d
  - Obtain the sum of these deviations i.e. $\sum d$
  - Apply the formula

$$\bar{X} = A + \frac{\sum d}{N}$$

Simple Arithmetic Mean – Individual Series

- Following are the monthly income of 10 employees in an office
  - 14780, 15760, 26690, 27750, 24840, 24920, 16100, 17810, 27050, 16950
- Calculate arithmetic mean of income

$\bar{X} = 21265$

$n = (3, 4, 7, 3, 5, 2, 6, 10)$

direct method

$$\bar{x} = \frac{40}{8} = 5$$

assumed mean

assumed mean $(A) = 10$

n	$d\ (\bar{x} - n)$
3	-7
4	-6
7	-3
3	-7
5	-5
2	-8
6	-4
10	0

$\Sigma d = \boxed{-40}$

$$\bar{x} = A + \frac{\Sigma d}{N}$$

$$= 10 + \frac{-40}{8}$$

$$= 10 - 5 = 5$$

Simple Arithmetic Mean – Discrete Series

- Direct method
- Steps
  - Multiply the frequency of each row with the variable and obtain the total $\sum fX$
  - Divide the total by number of observation that is the total frequency

Frequency value

$$\bar{X} = \frac{\sum fX}{N}$$

- Where
  - f = frequency
  - X = observations
  - N = total frequency = Σf

X No of Goals	f # students	f × x
0	8	0
1	10	10
2	12	24
3	3	9
4	5	20
5	2	10
	40	73

$73 = \Sigma fx$

$$\bar{x} = \frac{\Sigma fx}{N}$$

$$\bar{x} = \frac{73}{40} = \underline{\underline{1.82}}$$

Simple Arithmetic Mean – Discrete Series

- Shortcut method - Using Assumed mean
- Steps
  - Take an assumed mean and denote it by A
  - Take the deviations of the variable X from the assumed mean and denote the deviations by d
  - Multiply this deviation by respective frequency and take the total $\sum fd$
  - Apply the formula

$$\bar{X} = A + \frac{\sum fd}{N}$$

- Where
  - f = frequency
  - d = deviation from Assumed mean
  - A = assumed mean
  - N = total frequency

Simple Arithmetic Mean – Discrete Series

- From the following data of marks obtained by students, calculate arithmetic mean

Marks	20	30	40	50	60	70
# students	8	12	20	10	6	4

$$\bar{X} = \frac{\Sigma fx}{N} = \frac{2460}{60} = \underline{\underline{41}}$$

Simple Arithmetic Mean – Continuous Series

- Direct method
- Steps
  - Obtain the mid point of each class and denote it by m
  - Multiply these mid points by the respective frequency of each class and obtain $\sum fm$
  - Divide the total obtained by the sum of frequency (N)

mid point of range/class/bin

$$\bar{X} = \frac{\sum fm}{N}$$

- Where
  - f = frequency
  - m = mid point of each class
  - N = total frequency $= \sum f$

Simple Arithmetic Mean – Continuous Series

- Shortcut method - Using Assumed mean
- Steps
  - Take an assumed mean and denote it by A
  - From the mid point of each class deduct the assumed mean
  - Multiply the respective frequencies of each class by the deviations and obtain $\sum fd$
  - Apply formula

$$\bar{X} = A + \frac{\sum fd}{N}$$

- Where
  - f = frequency
  - d = deviation of class mid point from assumed mean
  - A = assumed mean
  - N = total frequency

x Daily Demand	f Frequency	m	fm
0-5	4	2.5	10.0
5-10	8	7.5	60.0
10-15	6	12.5	75.0
15-20	2	17.5	35.0
	20		180.0

$$\bar{x} = \frac{\Sigma fm}{N}$$

$$= \frac{180}{20}$$

$$\bar{x} = 9$$

Simple Arithmetic Mean – Continuous Series

- From the following data of marks obtained by students, calculate arithmetic mean

Marks	0-10	10-20	20-30	30-40	40-50	50-60
# students	5	10	25	30	20	10

$$\bar{X} = \frac{\Sigma fm}{N} = \frac{3300}{100} = \underline{\underline{33}}$$

Mathematical Properties of Arithmetic Mean

- Sum of the deviations of the items from the arithmetic mean (taking sign into account) is always zero
- Sum of the squared deviations of the items from arithmetic mean is minimum, that is, less than the sum of squared deviations of the items from any other value
- Including the mean value in the series multiple times wont change the mean ***
- If we have arithmetic mean and number of items of two or more than two related groups, we can compute combined mean of these groups using formula

$\bar{x} = \frac{15}{5} = 3$

x = 1 2 3 4 5

x	$x - \bar{x}$	$(x-\bar{x})^2$
1	-2	4
2	-1	1
3	0	0
4	1	1
5	2	4
	0	10

$$\overline{X_{12}} = \frac{N_1\overline{X_1} + N_2\overline{X_2}}{N_1 + N_2}$$

x_1 = 1 2 3 4 5

x_2 = 6 7 8 9 10

x_3 = 1, 2, 3, 4, 5, 6, 7, 8, 9, 10

$\bar{x}_1 = 5$

$\bar{x}_2 = 8$

$\bar{x}_3 = \frac{5\times5 + 5\times8}{5+8}$

$= \frac{25+40}{13} = \frac{65}{13} = 5$

x	$x - 5$	$(x-5)^2$
1	-4	16
2	-3	9
3	-2	4
4	-1	1
5	0	0

Merits

- It is simplest average to understand and easiest to compute
- It is affected by value of every item in the series
- It is defined by rigid mathematical formula with the result that everyone who computes the average gets the same answer
- It lends itself to subsequent algebraic treatment better than median or mode
- The mean is typical in the sense that it is the center of gravity, balancing the values on the either sides of it
- It is calculated values and not based on the positions

Geometric Mean

- Steps
  - Multiply all the values and get the result
  - Get the square root to the Nth power to find the geometric mean

$$\bar{X} = \sqrt[N]{x_1 * x_2 * \ldots .* x_n}$$

Harmonic Mean

- Steps
  - Get reciprocal of each number and add together
  - Divide the number of values by the total calculated eariler

$$\bar{X} = \frac{N}{\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_n}}$$

Weighted Mean

- Steps
  - Multiply every value with corresponding weight
  - Add the values together
  - Divide the total by sum of all the weights

$$\bar{X} = \frac{\sum WiXi}{W1 + W2 \ldots . + Wn}$$

$x = 1, 2, 3, 4, 5$

$\bar{x} = 3$

extreme value

Outlier

$\downarrow$

$x_2 = \underline{1 \quad 2 \quad 3 \quad 4 \quad 5} \quad 100$

$\bar{x}_2 = \frac{115}{6} = 12. . .$

position based average

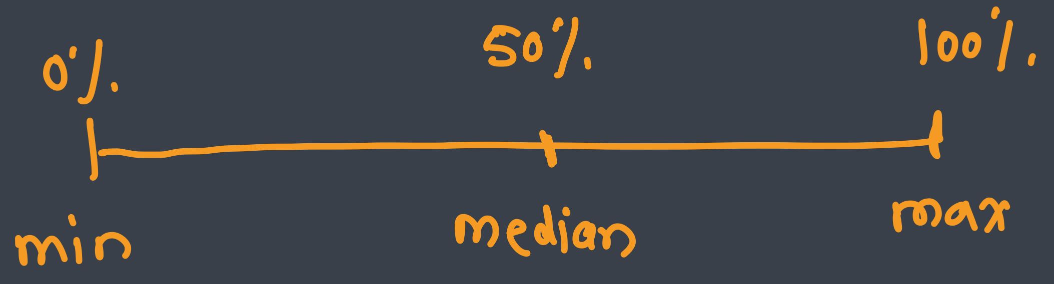
Median

middlemost value

Median

list of values

- By definition, it refers to the middle value in a distribution
- The median is just 50th percentile value below which 50% of the values in the sample fall
- It splits the observations into two halves
- Unlike the mean, median is calculated by position (which refers to the place of the value in the series)



ascending order

Median – Individual Series

- Steps
  - Arrange the data in the ascending or descending order of magnitude
  - In a group composed of an odd number of values such as 7, add 1 to the total number of values and divide it by 2. Thus 7 + 1 would be 8 which divided by 2 gives 4 – the position used to calculate the mean
  - In a group composed of even number of values such as 10, use the average of middle two values. Thus 10 / 2 gives 5 – which will produce a median by taking average of 5^{th} and 6^{th} position values

$$median = \frac{N+1}{2}$$

median

Odd no of values

$$\text{median} = \left(\frac{N+1}{2}\right)^{th} \text{value}$$

$N = 5$

$$\text{median} = \left(\frac{5+1}{2}\right) = 3^{rd} \text{ value}$$

20	25	30	45	60
1	2	3	4	5

Even no of values

median = average of two middle values

$$\text{median} = \text{avg}\left(\left(\frac{N}{2}\right)^{th}, \left(\frac{N}{2}+1\right)^{th}\right)$$

32.5

20	25	30	35	60	70
1	2	3	4	5	6

$$\text{position} = \left(\frac{6+1}{2}\right) = 3.5$$

$$\text{median} = \frac{30+35}{2} = 32.5$$

Median – Individual Series

- E.g. 1:
  - find median of 14100, 14150, 16080, 17120, 15200, 16160, 17400
  - Arrange them in ascending order
    - 14100, 14150, 15200, 16080, 16160, 17120, 17400
  - Median = (N + 1) / 2th item
  - Median = 7 + 1 / 2 = 4th item => 16080

- E.g. 2:
  - Find median of 19, 28, 40, 10, 29, 50, 37, 89, 90, 60
  - Arrange them in ascending order
    - 10, 19, 28, 29, 37, 40, 50, 60, 89, 90
  - Median = (N + 1)/ 2 the item
  - Median = average of 5th and 6th items => Average(37, 40) =>38.50

X	F
14100	1
14150	1
16080	1
17120	1
15200	1
16160	1
17400	1

Median – Discrete Series

- Steps
  - Arrange the data in ascending or descending order of magnitude
  - Find out cumulative frequencies
  - Apply the formula (N + 1) / 2 the item
  - Now look at the cumulative frequency and find the total which is either equal to (N + 1) /2 or next higher to that and determine the value of variable corresponding to it
  - This gives the value of median

$$\text{position} = \left(\frac{N+1}{2}\right)$$

Median – Discrete Series

20 20 20 20 . .

values

Marks	20	30	40	50	60	70
# students	8	12	20	10	6	4

frequency

$N = \Sigma f$

$N = 60$

median

Marks (x)	#students (f)	Cumulative frequency		
20	8	8	≥ 30.5	✗
30	12	20	≥ 30.5	✗
40	20	40	≥ 30.5	✓
50	10	50		
60	6	56		
70	4	60		
	60			

- Median is (N + 1) / 2 th item => (60 + 1) / 2 = 30.5 th item
- Since the value at 30.5th (or just higher than it) is 40
- Median = 40

No of Goals	# Students	C. Frequencies	
0	8	8 ≥ 20.5	✗
1	10	18 ≥ 20.5	✗
2	12	30 ≥ 20.5	✓
3	3	33	
4	5	38	
5	2	40	
	40		

$N = \Sigma f = 40$

$$\text{position} = \frac{N+1}{2} = \frac{41}{2}$$

$$\approx 20.5$$

median = 2

x	f
10	2
15	2
12	1
13	2
11	1

x	f	c.f
10	2	2
11	1	3
12	1	4
13	2	6 ≥ 4.5 ✓
15	2	8

$N = 8$, position $= \frac{8+1}{2} = 4.5$

median = 13

10 10 11 $\boxed{12 \ 13}$ 13 15 15

4

Median – Continuous Series

- Steps

median class

  - Determine the particular class in which the value of median lies, consider this as median class
  - Calculate the cumulative frequencies
  - Use N/2 as the rank of the median
  - Use the formula

$$median = L + \frac{\frac{N}{2} - cf}{f} * i$$

- Where
  - L = Lower limit of the median class (the class in which middle item of the distribution lies)
  - cf = cumulative frequency of the class preceding the median class
  - f = frequency of the median class
  - i = class interval of the median class

Median – Continuous Series

Marks	0-10	10-20	20-30	30-40	40-50	50-60
# students	5	10	25	30	20	10

Marks (x)	#students (f)	cf	
0-10	5	5	≥ 50 ✗
10-20	10	15	≥ 50 ✗
20-30	25	40	≥ 50 ✗
30-40	30	70	≥ 50 ✓
40-50	20	90	
50-60	10	100	

median class → 30-40

- The median class is => 100 /2 => 50 lies in (30-40)
- Median = 30 + ((100/2 – 40) / 30) * 10
- Median = 30 + (10/30) * 10 = 33.33

$N = \Sigma f = 100$

$\left(\frac{N}{2}\right) = \frac{100}{2} = 50$

$\text{median} = L + \frac{N/2 - cf}{f} \times i = 30 + \frac{50 - 40}{30} * 10 = 33.33$

x	f	c.f.
0-5	4	4 ← cf
median class → 5-10	8 ← f	12 ≥ 10 ✓
10-15	6	18
15-20	2	20

$N = \Sigma f = 20$

$\text{Position} = \frac{N}{2} = 10$

median class = $\underline{\underline{5-10}}$

$$\text{median} = L + \frac{N/2 - cF}{f} * i$$

$$= 5 + \frac{10 - 4}{8} * 5$$

$$\boxed{\text{median} = 8.75}$$

Merits

- It is useful in case of open-end classes since only the position and not the values of the items must be known
- Median is recommended if the distribution has unequal classes
- Extreme values do not affect the median as strongly as they do the mean
- It is most appropriate average in dealing with qualitative data
- Value of median can be calculated graphically
- It represents clear-cut the middle value in the distribution

names = [. . .]

Limitations

- For calculating median, it is necessary to arrange the data in a specific order
- Since it is a middle value, its value is not determined by each and every observation
- It is not capable of algebraic treatment
- The value of median is affected more by fluctuations than the value of the arithmetic mean
- It is erratic if the number of observations is very small

Mode

textual data

fix the NA/missing records

Mode = value with max frequency

- The mode or modal value is that value in a series which occurs most frequently
- That is the mode always will have the highest frequency in the data
- There are many situations where mean and median fails to reveal the true middle value, in such scenarios mode is used to find the central value

Mode – Individual Series

- Steps
  - Count the number of times the various values repeate themselves and the value occurring maximum number of times is the modal value

- E.g. 10, 28, 39, 40, 10, 20, 40, 50, 10 => mode = [10] → single mode
- E.g. 10, 20, 40, 50, 10, 20, 30, 40, 50 => mode = [10, 20, 50, 40] → multi mode
- E.g. 10, 20, 30, 40, 50, 60, 70, 80, 90 => mode = [] → No mode

Mode – Discrete Series

- Steps
  - Mode can be determined just be inspection
  - i.e. by looking to that value of the variable around which the items are most heavily concentrated

mode

- E.g.

Marks	20	30	40	50	60	70
# students	8	12	20	10	6	4

- The mode here is 40

Mode – Continuous Series

- Steps
  - Find the modal class by finding the largest value
  - Determine the value of mode by applying the following formula

$$mode = L + \frac{\Delta_1}{\Delta_1 + \Delta_2} * i$$

- Where
  - L = Lower limit of modal class
  - Δ_1 = difference between the frequency of modal class and frequency of pre-modal class
  - Δ_2 = difference between the frequency of modal class and frequency of post-modal class

  i = class interval

Mode – Continuous Series

modal class

Marks	0-10	10-20	20-30	30-40	40-50	50-60
# students	5	10	25	30	20	10

- Modal class here is: 30-40 → frequency 30 is max
- Using the formula
  - Mode = 30 + ((30-25) / ((30-25) + (30-20))) * 10
  - Mode = 30 + (5 / (5 + 10)) * 10
  - Mode = 30 + 3.33 = 33.33

$\Delta 1 = 30 - 25 = 5$

$\Delta 2 = 30 - 20 = 10$

Merits

- Mode is the most typical or representative value of the distribution
- Like median, mode is not unduly affected by extreme values
- It can be used to describe the qualitative phenomenon
- The value of mode can be calculated graphically

Limitations

- The value of mode can not always be determined
- It is not capable of algebraic manipulation
- The value of mode is not based on each and every value of distribution

10 12 13 14 14 15

10 - 20

① 12 15 13 11 10 5 9 2 5 → mean = 9.1, mode = 5, median = 10

② 3 8 10 5 2 9 3 1 → mean = 5.1 mode = 3 median = 4

③

x	f
5	2
3	4
2	5
1	6

mean = 2.23

mode = 1

median = 2

④

x	f
8	2
9	5
7	3
5	4

mean = 7.38

mode = 9

median = 8

⑤

x	f
5-10	5
10-15	9
15-20	6
20-25	3

mean = 14.02

mode = 12.85

median = 13.61

⑥

x	f
10-20	9
20-30	7
30-40	8
40-50	5
50-60	4

mean = 31.36

mode = 18.18

median = 30.62