

Chapter 14: Statistics



Statistics

1. Three measures of central tendency are:

- i. Mean
- ii. Median
- iii. Mode

2. The **arithmetic mean**, also called the average, is the quantity obtained by adding all the observations and then dividing by the total number of observations.

3. Arithmetic mean may be computed by anyone of the following methods:

- i. Direct method
- ii. Short-cut method/ Assumed mean method
- iii. Step-deviation method

4. **Direct method** of finding mean:

If a variant X takes values $x_1, x_2, x_3, \dots, x_n$ with corresponding frequencies $f_1, f_2, f_3, \dots, f_n$ respectively, then arithmetic mean of these values is given by:

$$\bar{X} = \frac{\sum_{i=1}^n f_i x_i}{N} \text{ where } N = \sum_{i=1}^n f_1 + f_2 + f_2 \dots \dots + f_n$$

5. **Class mark** = $\frac{1}{2}$ (Upper class limit + Lower class limit)

6. **Short-cut method/ assumed mean method** of finding mean:

Let $x_1, x_2, \dots, x_n$ be values of a variable X with corresponding frequencies $f_1, f_2, f_3, \dots, f_n$ respectively. Let A be the assumed mean. Then:

$$\bar{X} = A + \frac{1}{N} \left\{ \sum_{i=1}^n f_i d_i \right\}$$

Note that in case of continuous frequency distribution, the values of $x_1, x_2, x_3 \dots, x_n$ are taken as the mid-points or class-marks of the various classes.

7. **Step-deviation method** of finding mean:

Let $x_1, x_2, \dots, x_n$ be values of a variable X with corresponding frequencies $f_1, f_2, f_3, \dots, f_n$ respectively. Let A be the assumed mean. Then:

$$\bar{X} = A + h \left\{ \frac{1}{N} \sum_{i=1}^n f_i u_i \right\}$$

Here, h is generally taken as common factor of the deviations, in case of ungrouped frequency distribution. And, in case of grouped frequency distribution, h is the class

width, $u_i = \frac{x_i - A}{h} = \frac{d_i}{h}$

Note that in case of continuous frequency distribution, the values of $x_1, x_2, x_3 \dots, x_n$ are

taken as the mid-points or class-marks of the various classes.

8. The step deviation method will be convenient to apply if all the deviations (d 's) have a common factor.
9. If class mark obtained, are in decimal form, then step deviation method is preferred to calculate mean.
10. **Median** is a measure of central tendency which gives the value of the middle observation in the data, arranged in order. It is that value such that the number of observations above it is equal to the number of observations below it.

11. For finding the median of a raw data, we arrange the given data in increasing or decreasing order. If n is odd, then median is the value of $\left(\frac{n+1}{2}\right)^{th}$ observation.

If n is even, then median is the arithmetic mean of the values of $\left(\frac{n}{2}\right)^{th}$ and $\left(\frac{n}{2} + 1\right)^{th}$ observations.

12. The **cumulative frequency** of a class is the frequency obtained by adding the frequencies of all the classes preceding the given class to the frequency of the class.

13. In case of an **ungrouped frequency distribution**, we calculate the **median** by following the steps given below:

Step 1: Find the cumulative frequencies (c.f.) and obtain $N = \sum f_i$.

Step 2: Find $\frac{N}{2}$

Step 3: Look for the cumulative frequency (c. f.) just greater than $\frac{N}{2}$ and determine the corresponding value of the variable. The value so obtained is the median.

14. In case of a **continuous frequency distribution**, we calculate the **median** by following the steps:

Step 1: Find the cumulative frequencies (c.f.) and obtain $N = \sum f_i$.

Step 2: Find $\frac{N}{2}$

Step 3: Look for the cumulative frequency (c. f.) just greater than $\frac{N}{2}$ and determine the corresponding class. This class is known as the median class. (Note that the value of the median will lie in this class)

Step 4: Use the following formula to find median:

$$\text{Median} = l + \left[\frac{\frac{N}{2} - cf}{f} \right] \times h$$

Here, l = lower limit of the median class

f = frequency of the median class

h = width (size) of the median class

cf = cumulative frequency of the class preceding the median class

$$N = \sum f_1.$$

15. **Mode** is the value of the most frequently occurring observation in the data.
16. In an ungrouped frequency distribution, mode is the value of the variable having maximum frequency.
17. In a **grouped frequency distribution**, the modal class is the one with highest frequency and the **mode** can be calculated by the following formula

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

l = lower limit of the modal class

h = size of the class interval

f_1 = frequency of the modal class

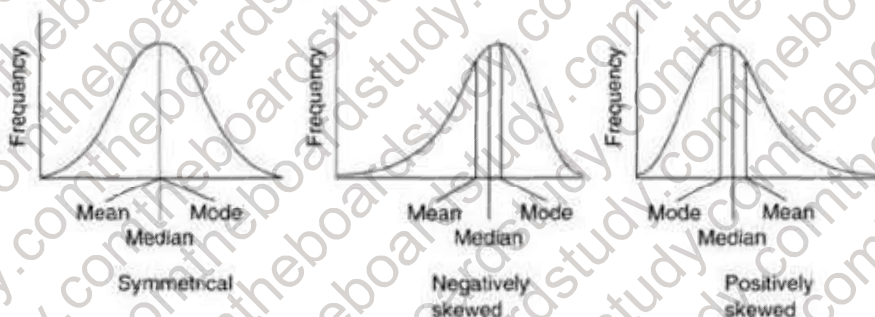
f_0 = frequency of the class preceding the modal class

f_2 = frequency of the class succeeding the modal class

18. The most frequently used measure of central tendency is the mean, because the mean is calculated by taking into account all the observations of a given data. And it lies between the smallest and the largest value of the data.
19. The biggest drawback in considering mean is that it is affected by the extreme values. One large or small number can distort the average. In that case, median is a better measure of central tendency. While, when the most repeated value or the most wanted one is required, then mode is used.
20. When all three measures of central tendency are equal, the distribution is called **symmetrical distribution**.
21. When the values of mean, median and mode are not equal, then the distribution is known as **asymmetrical or skewed**. In this case, the distribution can be positively skewed or negatively skewed.

Negatively skewed distributions have a few extremely low scores, while positively skewed distributions have a few extremely high scores.

- When the data is negatively skewed, then $\text{Mean} < \text{Median} < \text{Mode}$
- When the positively skewed, then $\text{Mean} > \text{Median} > \text{Mode}$



Three measure of central values are connected by the following relation:

$$3 \text{ Median} = \text{Mode} + 2 \text{ Mean}$$

22. The **cumulative frequency** is the accumulated or sum of frequencies up to a particular point. A table showing the cumulative frequencies is called a **cumulative frequency distribution**.

23. There are two types of cumulative frequencies:

- i. **Less than type cumulative frequency distribution:** It is found by adding sequentially the frequencies of all the earlier classes including the class adjacent to which it is written. The cumulate is started from the lowest to the highest size.
- ii. **More than type cumulative frequency distribution:** It is obtained by finding the cumulate of frequencies starting from the highest to the lowest class.

24. A cumulative frequency distribution can be represented graphically by means of an **ogive**.

25. There are two types of ogives:

- i. **'Less than' ogive:** In a less than ogive the upper limit of a class (x axis) is plotted against its cumulative frequency (y axis) as a point on the ogive. The 'less than ogive' is a rising curve.
- ii. **'More than' ogive:** In a 'more than ogive' the lower limit of a class (x axis) is plotted against its cumulative frequency (y axis) as a point on the ogive. The 'more than ogive' is a falling curve.

26. The ogives can be drawn only when the given class intervals are continuous and if this is not the case then first the class intervals are made continuous.

27. In order to determine the **median from less than ogive or more than ogive**, we follow the steps given below:

Step 1: Draw more than or less than ogive as asked in question. Find of observations, $\frac{N}{2}$ where N is the total number

Step 2: Locate the $\frac{N}{2}$ cumulative frequency on the y-axis.

Step 3: Draw a line parallel to x-axis through the point obtained in step 2, cutting the cumulative frequency curve at a point P (say).

Step 4: Draw perpendicular PM from P on the x-axis. The x-coordinate of point M is the median value.

28. If we draw **less than ogive and more than ogive on the same graph**, then median can be obtained by following the steps given below:

Step 1: Draw both ogives on the same graph.

Step 2: Identify the point of intersection of both ogives and mark it as Q (say).

Step 3: Draw perpendicular from Q on x-axis.

Step 4: The point of perpendicular on x-axis is the median.

Ungrouped Data

Ungrouped data is data in its original or raw form. The observations are not classified into groups.

For example, the ages of everyone present in a classroom of kindergarten kids with the teacher is as follows:

3, 3, 4, 3, 5, 4, 3, 3, 4, 3, 3, 3, 3, 4, 3, 27.

This data shows that there is one adult present in this class and that is the teacher.

Ungrouped data is easy to work with when the data set is small.

Grouped Data

In grouped data, observations are organized in groups.

For example, a class of students got different marks in a school exam. The data is tabulated as follows:

Mark interval	0-20	21-40	41-60	61-80	81-100
No. of Students	13	9	36	32	10

This shows how many students got the particular mark range. Grouped data is easier to work with when a large amount of data is present.

Frequency

Frequency is the number of times a particular observation occurs in data.

Class Interval

Data can be grouped into class intervals such that all observations in that range belong to that class.

Class width = upper class limit – lower class limit

Mean

Finding the mean for Grouped Data when class Intervals are not given

For grouped data without class intervals,

$$\text{Mean} = \bar{x} = \frac{\sum x_i f_i}{\sum f_i}$$

where f_i is the frequency of i^{th} observation x_i .

Finding the mean for Grouped Data when class Intervals are given

For grouped data with class intervals,

$$\text{Mean} = \bar{x} = \frac{\sum x_i f_i}{\sum f_i}$$

Where f_i is the frequency of i^{th} class whose class mark is x_i .

Classmark = (Upper Class Limit+ Lower Class Limit)/2

Direct method of finding mean

Step 1: Classify the data into intervals and find the corresponding frequency of each class.

Step 2: Find the class mark by taking the midpoint of the upper and lower class limits.

Step 3: Tabulate the product of the class mark and its corresponding frequency for each class. Calculate their sum ($\sum x_i f_i$).

Step 4: Divide the above sum by the sum of frequencies ($\sum f_i$) to get the mean.

Assumed mean method of finding mean

Step 1: Classify the data into intervals and find the corresponding frequency of each class.

Step 2: Find the class mark by taking the midpoint of the upper and lower class limits.

Step 3: Take one of the x_i 's (usually one in the middle) as the assumed mean and denote it by 'a'.

Step 4: Find the deviation of 'a' from each of the x_i 's

$$d_i = x_i - a$$

Step 5: Find the mean of the deviations

$$\bar{d} = \frac{\sum f_i d_i}{\sum f_i}$$

Step 6: Calculate the mean as

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

The relation between the Mean of deviations and mean

$$d_i = x_i - a$$

Summing over all x_i 's,

$$\sum d_i = \sum x_i - \sum a$$

Dividing throughout by $\sum f_i = n$, Where 'n' is the total number of observations.

$$\bar{d} = \bar{x} - a$$

Step-Deviation method of finding mean

Step 1: Classify the data into intervals and find the corresponding frequency of each class.

Step 2: Find the class mark by taking the midpoint of the upper and lower class limits.

Step 3: Take one of the x_i 's (usually one in the middle) as assumed mean and denote it by 'a'.

Step 4: Find the deviation of a from each of the x_i 's

$$d_i = x_i - a$$

Step 5: Divide all deviations $-d_i$ by the class width (h) to get u_i 's.

$$u_i = \frac{x_i - a}{h}$$

Step 6: Find the mean of u_i s

$$\bar{u} = \frac{\sum f_i u_i}{\sum f_i}$$

Step 7: Calculate the mean as

$$\bar{x} = a + h \times \frac{\sum f_i u_i}{\sum f_i} = a + h\bar{u}$$

Relation between mean of Step- Deviations (u) and mean

$$u_i = \frac{x_i - a}{h}$$

$$\bar{u} = \frac{\sum f_i \frac{x_i - a}{h}}{\sum f_i}$$

$$\bar{u} = \frac{1}{h} \times \frac{\sum f_i x_i - a \sum f_i}{\sum f_i}$$

$$\bar{u} = \frac{1}{h} \times (\bar{x} - a)$$

Important relations between methods of finding mean

- All three methods of finding mean yield the same result.
- Step deviation method is easier to apply if all the deviations have a common factor.
- Assumed mean method and step deviation method are simplified versions of the direct method.

Median

Finding the Median of Grouped Data when class Intervals are not given

Step 1: Tabulate the observations and the corresponding frequency in ascending or descending order.

Step 2: Add the cumulative frequency column to the table by finding the cumulative frequency up to each observation.

Step 3: If the number of observations is odd, the median is the observation whose cumulative frequency is just greater than or equal to $(n+1)/2$.

If the number of observations is even, the median is the average of observations whose cumulative frequency is just greater than or equal to $n/2$ and $(n/2)+1$.

Cumulative Frequency

Cumulative frequency is obtained by adding all the frequencies up to a certain point.

Finding median for Grouped Data when class Intervals are given

Step 1: find the cumulative frequency for all class intervals.

Step 2: the median class is the class whose cumulative frequency is greater than or nearest to $n/2$, where n is the number of observations.

Step 3: Median = $l + [(N/2 - cf)/f] \times h$

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 (assuming class size to be equal).

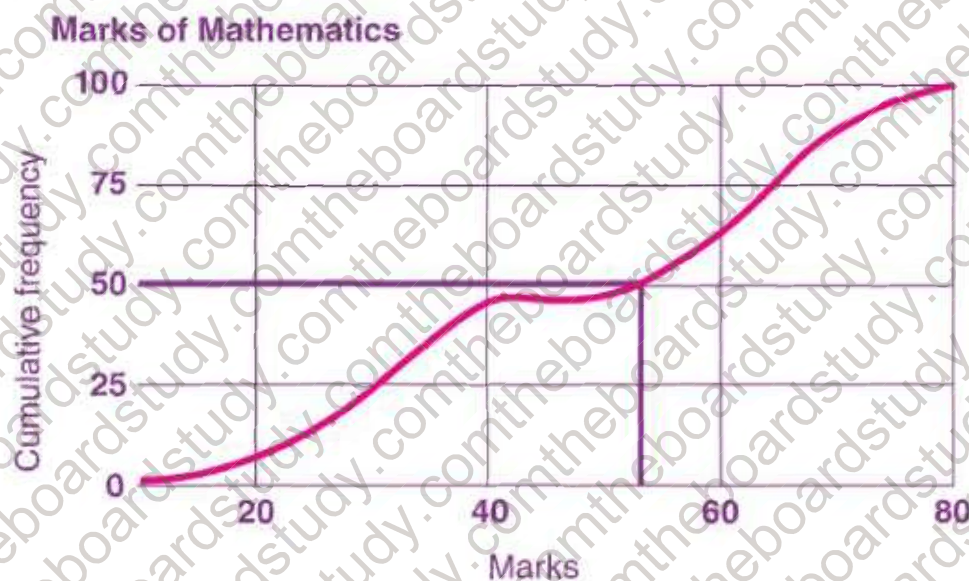
Cumulative Frequency distribution of less than type

Cumulative frequency of the less than type indicates the number of observations which are less than or equal to a particular observation.

Cumulative Frequency distribution of more than type

Cumulative frequency of more than type indicates the number of observations that are greater than or equal to a particular observation.

Visualising formula for median graphically



Median from Cumulative Frequency Curve

Step 1: Identify the median class.

Step 2: Mark cumulative frequencies on the y-axis and observations on the x-axis corresponding to the median class.

Step 3: Draw a straight line graph joining the extremes of class and cumulative frequencies.

Step 4: Identify the point on the graph corresponding to $cf = n/2$

Step 5: Drop a perpendicular from this point onto the x-axis.

Ogive of less than type

The graph of a cumulative frequency distribution of the less than type is called an 'ogive of the less than type'.

Ogive of more than type

The graph of a cumulative frequency distribution of the more than type is called an 'ogive of the more than type'.

Relation between the less than and more than type curves

The point of intersection of the ogives of more than and less than types gives the median of the grouped frequency distribution.

Mode

Finding mode for Grouped Data when class intervals are not given

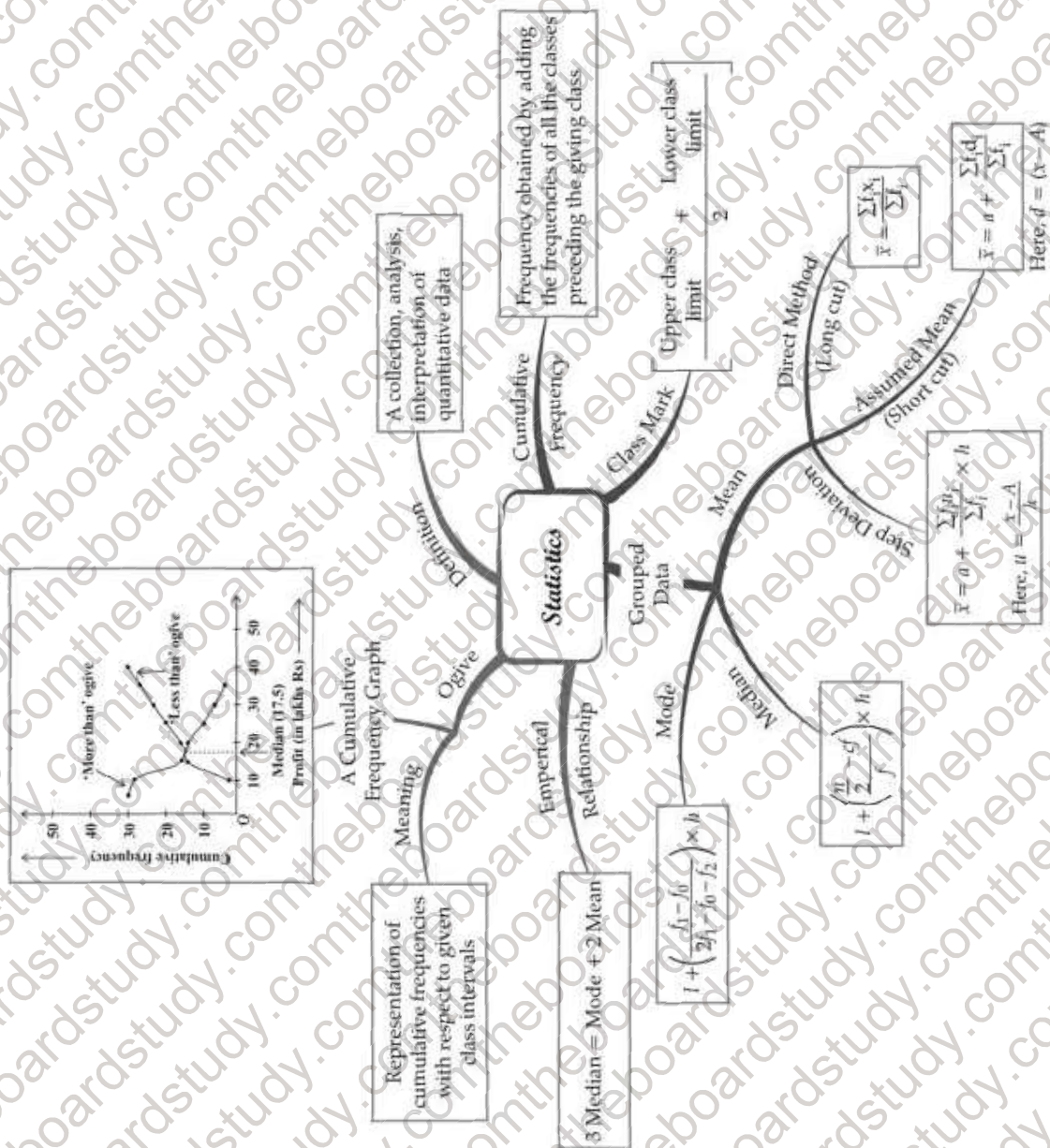
In grouped data without class intervals, the observation having the largest frequency is the mode.

Finding mode for Ungrouped Data

For ungrouped data, the mode can be found out by counting the observations and using tally marks to construct a frequency table.

The observation having the largest frequency is the mode.

MIND MAP : LEARNING MADE SIMPLE



Important Questions

Multiple Choice questions-

1. Cumulative frequency curve is also called

- (a) histogram
- (b) ogive
- (c) bar graph
- (d) median

2. The relationship between mean, median and mode for a moderately skewed distribution is

- (a) $\text{mode} = \text{median} - 2 \text{ mean}$
- (b) $\text{mode} = 3 \text{ median} - 2 \text{ mean}$
- (c) $\text{mode} = 2 \text{ median} - 3 \text{ mean}$
- (d) $\text{mode} = \text{median} - \text{mean}$

3. The median of set of 9 distinct observations is 20.5. If each of the largest 4 observations of the set is increased by 2, then the median of the new set

- (a) is increased by 2
- (b) is decreased by 2
- (c) is two times of the original number
- (d) Remains the same as that of the original set.

4. Mode and mean of a data are 12k and 15A. Median of the data is

- (a) 12k
- (b) 14k
- (c) 15k
- (d) 16k

5. The times, in seconds, taken by 150 athletes to run a 110 m hurdle race are tabulated below:

Class	Frequency
13.8 – 14.0	2
14.0 – 14.2	4
14.2 – 14.4	5
14.4 – 14.6	71
14.6 – 14.8	48
14.8 – 15.0	20

The number of athletes who completed the race in less than 14.6 seconds is:

- (a) 11
- (b) 71
- (c) 82
- (d) 130

6. The abscissa of the point of intersection of the less than type and of the more than type cumulative frequency curves of a grouped data gives its

- (a) mean
- (b) median
- (c) mode
- (d) all the three above

7. While computing mean of grouped data, we assume that the frequencies are:

- (a) evenly distributed over all the classes
- (b) centred at the classmarks of the classes
- (c) centred at the upper limits of the classes
- (d) centred at the lower limits of the classes

8. Mean of 100 items is 49. It was discovered that three items which should have been 60, 70, 80 were wrongly read as 40, 20, 50 respectively. The correct mean is

(a) 48

(b) 49

(c) 50

(d) 60

9. While computing mean of grouped data, we assume that the frequencies are

(a) centred at the upper limits of the classes

(b) centred at the lower limits of the classes

(c) centred at the classmarks of the classes

(d) evenly distributed over all the classes

10. Which of the following can not be determined graphically?

(a) Mean

(b) Median

(c) Mode

(d) None of these

Very Short Questions:

1. In a continuous frequency distribution, the median of the data is 21. If each observation is increased by 5, then find the new median.

2. From the following frequency distribution, find the median class:

Cost of living index	No. of weeks
1400-1550	8
1550-1700	15
1700-1850	21
1850-2000	8

3. Consider the following distribution, find the frequency of class 30-40.

Marks obtained	No. of Students
0 or more	63
10 or more	58
20 or more	55
30 or more	51
40 or more	48
50 or more	42

4. Following table shows sale of shoes in a store during one month:

Size of shoe	No. of pairs sold
3	4
4	18
5	25
6	12
7	5
8	1

Find the modal size of the shoes sold.

5. Weekly household expenditure of families living in a housing society are shown below:

Weekly expenditure (in ₹)	No. of families (f)
Up to 3000	4
3000-6000	25
6000-9000	31
9000-12000	48
12000-15000	10

Find the upper limit of the modal class.

- Find the class mark of the class 10 – 25.
- Find the mean of the first five natural numbers.
- A data has 13 observations arranged in descending order. Which observation represents the median of data?
- If the mode of a distribution is 8 and its mean is also 8, then find median.
- In an arranged series of an even number of $2n$ terms which term is median?

Short Questions :

- If x_i 's are the mid-points of the class intervals of a grouped data. f_i 's are the

corresponding frequencies and is the mean, then find $\sum f_i (x_i - \bar{x})$.

2. Consider the following frequency distribution.

Class	0-5	6-11	12-17	18-23	24-29
Frequency	13	10	15	8	11

3. Find the median class of the following distribution:

Class	0-10	10-20	20-30	30-40	40-50	50-60	60-70
Frequency	4	4	8	10	12	8	4

4. Find the class marks of classes 15.5 – 18.5 and 50 – 75.

5. If the mean of the following distribution is 6, find the value of p.

x	2	4	6	10	p + 5
f	3	2	3	1	2

6. Find the mean of the following distribution:

x	4	6	9	10	15
f	5	10	10	7	8

7. The following data gives the information on the observed lifetimes (in hours) of 225 electrical components:

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

Determine the modal lifetimes of the components.

8. The distribution below gives the weights of 30 students of a class. Find the median weight of the students.

Weight (in kg)	40-45	45-50	50-55	55-60	60-65	65-70	70-75
Number of students	2	3	8	6	6	3	2

Long Questions :

1. The following table gives the literacy rate (in percentage) of 35 cities. Find the mean literacy rate.

Literacy rate (in %)	45-55	55-65	65-75	75-85	85-90
Number of Cities	3	10	11	8	3

2. The following distribution shows the daily pocket allowance of children of a locality. The mean pocket allowance is ₹ 18. Find the missing frequency f .

Daily pocket allowance (in ₹)	11-13	13-15	15-17	17-19	19-21	21-23	23-25
Number of children	7	6	9	13	f	5	4

3. The mean of the following frequency distribution is 62.8. Find the missing frequency x .

Classes	0-20	20-40	40-60	60-80	80-100	100-120
Frequency	5	8	x	12	7	8

4. The distribution below gives the marks of 100 students of a class.

Marks	0-5	5-10	10-15	15-20	20-25	25-30	30-35	35-40
Number of students	4	6	10	10	25	22	18	5

5. During the medical check-up of 35 students of a class, their weights were recorded as follows:

Weight (in kg)	Number of students	Weight (in kg)	Number of students
Less than 38	0	Less than 46	14
Less than 40	3	Less than 48	28
Less than 42	5	Less than 50	32
Less than 44	9	Less than 52	35

Draw a less than type ogive for the given data. Hence, obtain the median weight from the graph and verify the result by using the formula.

Case Study Questions:

1. A petrol pump owner wants to analyse the daily need of diesel at the pump. For this he collected the data of vehicles visited in 1hr. The following frequency distribution table shows the classification of the number of vehicles and quantity of diesel filled in them.

Diesel Filled (in Litres)	3-5	5-7	7-9	9-11	11-13
Number of vehicles	5	10	10	7	8



i. Which of the following is correct?

- If x_i and f_i are sufficiently small, then direct method is appropriate choice for calculating mean.
- If x_i and f_i are sufficiently large, then direct method is appropriate choice for calculating mean.
- If x_i and f_i are sufficiently small, then assumed mean method is appropriate choice for calculating mean.
- None of the above.

ii. Average diesel required for a vehicle is:

- 8.15 litres
- 6 litres
- 7 litres
- 5.5 litres

iii. If approximately 2000 vehicles comes daily at the petrol pump, then how much litres of diesel the pump should have?

- 16200 litres
- 16300 litres
- 10600 litres

d. 15000litres

iv. The sum of upper and lower limit of median class is:

a. 22

b. 10

c. 16

d. None of this.

v. If the median of given data is 8 litres, then mode will be equal to:

a. 7.5 litres

b. 7.7 litres

c. 5.7 litres

d. 8 litres

2. A bread manufacturer wants to know the lifetime of the product. For this, he tested the lifetime of 400 packets of bread. The following tables gives the distribution of the lifetime of 400 packets.

Lifetime (in hours)	Number of packets (Cumulative frequency)
150-200	14
200-250	70
250-300	130
300-350	216
350-400	290
400-450	352
450-500	400

i. If m be the class mark and b be the upper limit of a class in a continuous frequency distribution, then lower limit of the class is:

a. $2m + \sqrt{b}$

b. $2m + b$

c. $m - b$

d. $2m - b$

ii. The average lifetime of a packet is:

a. 341hrs

- b. 300hrs
c. 340hrs
d. 301hrs
- iii. The median lifetime of a packet is:
a. 347hrs
b. 340hrs
c. 346hrs
d. 342hrs
- iv. If empirical formula is used, then modal lifetime of a packet is:
a. 340hrs
b. 341hrs
c. 348hrs
d. 349hrs
- v. Manufacturer should claim that the lifetime of a packet is:
a. 346hrs
b. 341hrs
c. 340hrs
d. 347hrs

Assertion Reason Questions-

1. **Directions:** In the following questions, a statement of assertion (A) is followed by a statement of reason (R). Mark the correct choice as:

- (a) Both A and R are true and R is the correct explanation of A.
(b) Both A and R are true and R is not the correct explanation of A.
(c) A is true but R is false.
(d) Both A and R is false.

Assertion: median = $\frac{(n+1)/2}{}^{\text{th}}$ value if n is odd

Reason: If the number of runs scored by 11 players of a cricket team of India are 5, 19, 42, 11, 50, 30, 21, 0, 52, 36, 27 then median is 30

2. **Directions:** In the following questions, a statement of assertion (A) is followed by a statement of reason (R). Mark the correct choice as:

- Both A and R are true and R is the correct explanation of A.
- Both A and R are true and R is not the correct explanation of A.
- A is true but R is false.
- Both A and R is false.

Assertion: if the value of mode and mean is 60 and 66 then the value of median is 64.

Reason: median = (mode + 2mean)

Answer Key-

Multiple Choice questions-

- (b) ogive
- (b) mode = 3 median – 2 mean
- (d) Remains the same as that of the original set.
- (b) 14k
- (c) 82
- (b) median
- (b) centred at the classmarks of the classes
- (c) 50
- (c) centred at the classmarks of the classes
- (a) Mean

Very Short Answer :

- New median = $21 + 5 = 26$
-

Cost of living index	No. of weeks (<i>f</i>)	c.f.
1400-1550	8	8
1550-1700	15	23
1700-1850	21	44
1850-2000	8	52
	52	

Here, $n = 52$, $\frac{n}{2} = \frac{52}{2} = 26$

∴ Median class 1700 – 1850.

3.

Marks obtained	No. of students (c.f.)	Class Interval	<i>f</i>
0 or more	63	0-10	5
10 or more	58	10-20	3
20 or more	55	20-30	4
30 or more	51	30-40	3
40 or more	48	40-50	6
50 or more	42	50-60	42
			63

∴ Frequency of class 30 – 40 = 3

4. Maximum no. of pairs sold = 25 (size 5)

∴ Modal size of shoes = 5

5. Maximum frequency = 48

∴ Modal class = 9,000 – 12,000

Upper limit of the modal class = 12,000

6.

$$\text{Class mark} = \frac{\text{Upper limit} + \text{Lower limit}}{2} = \frac{10 + 25}{2} = \frac{35}{2} = 17.5$$

7.

$$\text{Mean} = \frac{x_1 + x_2 + x_3 + x_4 + x_5}{5} = \frac{1 + 2 + 3 + 4 + 5}{5} = \frac{15}{5} = 3$$

8. Total no. of observations = 13, which is odd

$$\therefore \text{The median will be } \left(\frac{n+1}{2}\right)^{\text{th}} \text{ term} = \left(\frac{13+1}{2}\right)^{\text{th}} = \left(\frac{14}{2}\right)^{\text{th}} = 7^{\text{th}}$$

i.e., 7th term will be the median.

9. Mode = 8; Mean = 8; Median = ?

Relation among mean, median and mode is

$$3 \text{ median} = \text{mode} + 2 \text{ mean}$$

$$3 \times \text{median} = 8 + 2 \times 8$$

$$\text{Median} = \frac{8+16}{3} = \frac{24}{3} = 8$$

10. No. of terms = 2n which are even

$$\therefore \text{The median term will be } \frac{\left[\left(\frac{n}{2}\right)^{\text{th}} + \left(\frac{n}{2} + 1\right)^{\text{th}}\right]}{2}$$

$$\text{Put } n = 2n$$

$$= \frac{\left[\left(\frac{2n}{2}\right)^{\text{th}} + \left(\frac{2n}{2} + 1\right)^{\text{th}}\right]}{2} = \left[\frac{n^{\text{th}} + (n+1)^{\text{th}}}{2}\right]$$

i.e., the mean of nth and (n + 1)th term will be the median.

Short Answer :

1.

$$\text{We know mean } (\bar{x}) = \frac{\sum f_i x_i}{\sum f_i}$$

$$\therefore \sum f_i x_i = \bar{x} \sum f_i \quad \dots (i)$$

$$\text{Now the value of } \sum f_i (x_i - \bar{x}) = \sum f_i x_i - \sum f_i \bar{x}$$

$$= \sum f_i \bar{x} - \sum f_i \bar{x} = 0 \quad [\text{Using (i)}]$$

2. Classes are not continuous, hence make them continuous by adding 0.5 to the upper limits and subtracting 0.5 from the lower limits.

C.I.	Frequency	Cumulative Frequency
0-5.5	13	13
5.5-11.5	10	23
11.5-17.5	15	38
17.5-23.5	08	46
23.5-29.5	11	57
Total	$\Sigma f = 57$	

Class interval can't be negative hence the first CI is starting from 0.

Now to find median class we calculate $\frac{\Sigma f}{2} = \frac{57}{2} = 28.5$

$\therefore$ Median class = 11.5 – 17.5.

So, the upper limit is 17.5

3. First we find the cumulative frequency

Classes	Frequency	Cumulative Frequency
0-10	4	4
10-20	4	8
20-30	8	16
30-40	10	26
40-50	12	38
50-60	8	46
60-70	4	50
Total	50	

Here, $\frac{n}{2} = \frac{50}{2}$

$\therefore$ Median class = 30 – 40.

- 4.

Class marks = $\frac{\text{upper limit} + \text{lower limit}}{2}$

$$\text{Class marks of } 15.5 - 18.5 = \frac{18.5 + 15.5}{2} = \frac{34}{2} = 17$$

$$\text{Class marks of } 50 - 75 = \frac{75 + 50}{2} = \frac{125}{2} = 62.5$$

5. Calculation of mean

x_i	f_i	$f_i x_i$
2	3	6
4	2	8
6	3	18
10	1	10
$p + 5$	2	$2p + 10$
Total	$\Sigma f_i = 11$	$\Sigma f_i x_i = 2p + 52$

We have, $\Sigma f_i = 11$, $\Sigma f_i x_i = 2p + 52$, $\bar{X} = 6$.

$$\begin{aligned} \therefore \text{Mean } (\bar{X}) &= \frac{\Sigma f_i x_i}{\Sigma f_i} \\ \Rightarrow 6 &= \frac{2p + 52}{11} \Rightarrow 66 = 2p + 52 \\ \Rightarrow 2p &= 14 \Rightarrow p = 7 \end{aligned}$$

6. Calculation of arithmetic mean

x_i	f_i	$f_i x_i$
4	5	20
6	10	60
9	10	90
10	7	70
15	8	120
Total	$\Sigma f_i = 40$	$\Sigma f_i x_i = 360$

$$\therefore \text{Mean } (\bar{X}) = \frac{\Sigma f_i x_i}{\Sigma f_i} = \frac{360}{40} = 9$$

7. Here, the maximum class frequency is 61 and the class corresponding to this frequency is 60 – 80.

So, the modal class is 60 – 80.

Here, $l = 60$, $h = 20$, $f_1 = 61$, $f_0 = 52$, $f_2 = 38$

$$\begin{aligned} \therefore \text{Mode} &= l + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times h = 60 + \frac{61 - 52}{2 \times 61 - 52 - 38} \times 20 = 60 + \frac{9}{122 - 90} \times 20 \\ &= 60 + \frac{9}{32} \times 20 = 60 + \frac{45}{8} = 60 + 5.625 = 65.625 \end{aligned}$$

Hence, modal lifetime of the components is 65.625 hours.

8. Calculation of median

Weight (in kg)	Number of students (f)	Cumulative frequency (cf)
40-45	2	2
45-50	3	5
50-55	8	13
55-60	6	19
60-65	6	25
65-70	3	28
70-75	2	30
Total	$\Sigma f_i = 30$	

The cumulative frequency just greater than $\frac{n}{2} = 15$ is 19, and the corresponding class is 55 – 60.

$\therefore$ 55 – 60 is the median class.

$$\begin{aligned} \therefore \text{Median} &= l + \left(\frac{\frac{n}{2} - cf}{f} \right) \times h \\ &= 55 + \left(\frac{15 - 13}{6} \right) \times 5 = 55 + \frac{2}{6} \times 5 = 55 + 1.67 = 56.67 \end{aligned}$$

Hence, median weight is 56.67 kg.

Long Answer :

- Here, we use step deviation method to find mean.

Let assumed mean $A = 70$ and class size $h = 10$

$$\text{So, } u_i = \frac{x_i - 70}{10}$$

Now, we have

Literacy rate (in %)	Frequency	Class mark	$u_i = \frac{x_i - 70}{10}$	$f_i u_i$
45-55	3	50	-2	-6
55-65	10	60	-1	-10
65-75	11	70	0	0
75-85	8	80	1	8
85-95	3	90	2	6
Total	$\Sigma f_i = 35$			$\Sigma f_i u_i = -2$

$$\therefore \text{Mean } (\bar{X}) = A + h \times \frac{\Sigma f_i u_i}{\Sigma f_i} = 70 + 10 \times \frac{(-2)}{35} = 70 - 0.57 = 69.43\%$$

2. Let the assumed mean $A = 16$ and class size $h = 2$, here we apply step deviation method.

$$\text{So, } u_i = \frac{x_i - A}{h} = \frac{x_i - 16}{2}$$

Now, we have,

Class interval	Frequency	Class mark	$u_i = \frac{x_i - 16}{2}$	$f_i u_i$
11-13	7	12	-2	-14
13-15	6	14	-1	-6
15-17	9	16	0	0
17-19	13	18	1	13
19-21	f	20	2	$2f$
21-23	5	22	3	15
23-25	4	24	4	16
Total	$\Sigma f_i = f + 44$			$\Sigma f_i u_i = 2f + 24$

We have, Mean $(\bar{X}) = 18$, $A = 16$ and $h = 2$

$$\therefore \bar{X} = A + h \times \frac{\Sigma f_i u_i}{\Sigma f_i}$$

$$18 = 16 + 2 \times \left(\frac{2f + 24}{f + 44} \right) \Rightarrow 2 = 2 \times \left(\frac{2f + 24}{f + 44} \right)$$

$$\Rightarrow 1 = \frac{2f + 24}{f + 44}$$

$$\Rightarrow f = 44 - 24$$

$$\Rightarrow f = 20$$

$$\Rightarrow f + 44 = 2f + 24$$

Hence, the missing frequency is 20.

3. We have

Class interval	Frequency (f_i)	Class mark (x_i)	$f_i x_i$
0-20	5	10	50
20-40	8	30	240
40-60	x	50	$50x$
60-80	12	70	840
80-100	7	90	630
100-120	8	110	880
Total	$\Sigma f_i = 40 + x$		$\Sigma f_i x_i = 2640 + 50x$

Here, $\Sigma f_i x_i = 2640 + 50x$, $\Sigma f_i = 40 + x$, $\bar{X} = 62.8$

$$\therefore \text{Mean } (\bar{X}) = \frac{\Sigma f_i x_i}{\Sigma f_i}$$

$$\Rightarrow 62.8 = \frac{2640 + 50x}{40 + x}$$

$$\Rightarrow 2512 + 62.8x = 2640 + 50x$$

$$\Rightarrow 62.8x - 50x = 2640 - 2512$$

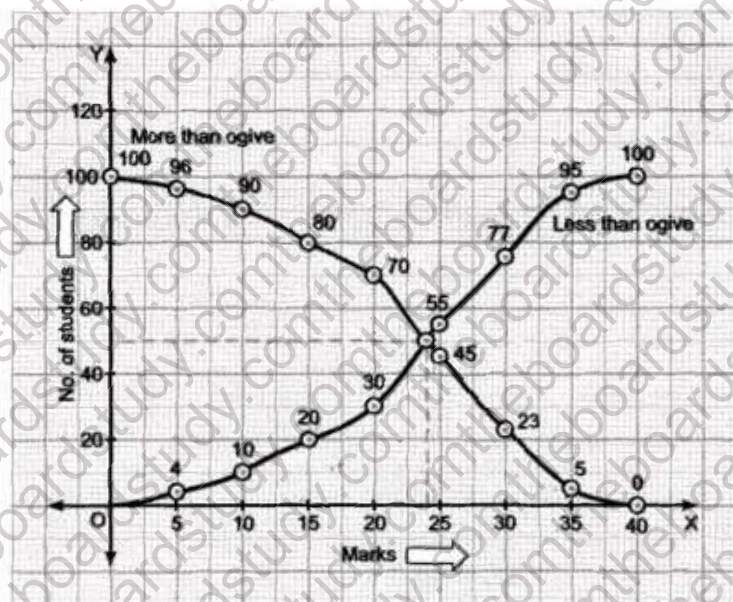
$$\Rightarrow 12.8x = 128$$

$$\therefore x = \frac{128}{12.8} = 10$$

Hence, the missing frequency is 10.

4.

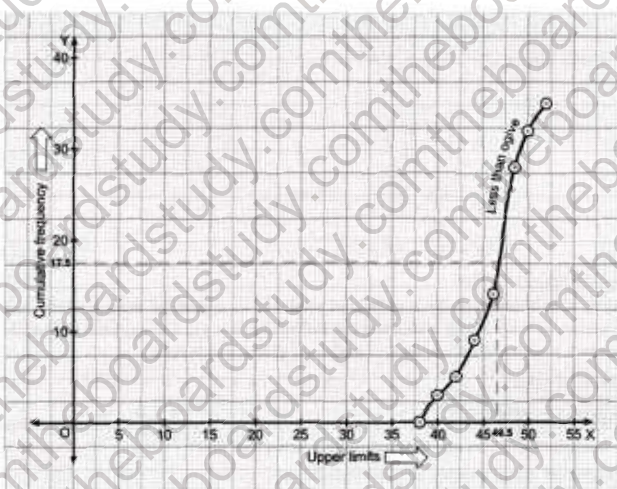
Marks	Cumulative Frequency	Marks	Cumulative Frequency
Less than 5	4	More than 0	100
Less than 10	10	More than 5	96
Less than 15	20	More than 10	90
Less than 20	30	More than 15	80
Less than 25	55	More than 20	70
Less than 30	77	More than 25	45
Less than 35	95	More than 30	23
Less than 40	100	More than 35	5



Hence, median marks = 24

- To represent the data in the table graphically, we mark the upper limits of the class interval on x-axis and their corresponding cumulative frequency on y-axis choosing a convenient scale. Now, let us plot the points corresponding to the ordered pair given by (38,0), (40,3), (42,5), (44, 9), (46, 14), (48, 28), (50, 32) and (52, 35) on a graph paper and join them by a freehand smooth curve.

Thus, the curve obtained is the less than type ogive.



Now, locate $\frac{n}{2} = \frac{35}{2} = 17.5$ on the y-axis,

We draw a line from this point parallel to x-axis cutting the curve at a point. From this point, draw a perpendicular line to the x-axis. The point of intersection of this perpendicular with the x-axis gives the median of the data.

Here it is 46.5.

Let us make the following table in order to find median by using formula.

Weight (in kg)	No. of Students (frequency) (f_i)	Cumulative frequency (cf)
36-38	0	0
38-40	3	3
40-42	2	5
42-44	4	9
44-46	5	14
46-48	14	28
48-50	4	32
50-52	3	35
Total	$\Sigma f_i = 35$	

Here, $n = 35$, $\frac{n}{2} = \frac{35}{2} = 17.5$, cumulative frequency greater than $\frac{n}{2} = 17.5$ is 28 and corresponding class is 46 – 48. So median class is 46 – 48.

Now, we have $l = 46$, $\frac{n}{2} = 17.5$, $cf = 14$, $f = 14$, $h = 2$

$$\begin{aligned}
 \therefore \text{Median} &= l + \left(\frac{\frac{n}{2} - cf}{f} \right) \times h \\
 &= 46 + \left(\frac{17.5 - 14}{14} \right) \times 2 \\
 &= 46 + \frac{3.5}{14} \times 2 = 46 + \frac{7}{14} \\
 &= 46 + 0.5 = 46.5
 \end{aligned}$$

Hence, median is verified.

Case Study Answer-

1. Answer:

- i. (a) If x_i and f_i are sufficiently large, then direct method is appropriate choice for calculating mean.

Solution:

If f_i and x_i are very small, then direct method is appropriate method for calculating mean.

- ii. (a) 8.15 litres

Solution:

The frequency distribution table from the given data can be drawn as:

Class	Class mark (x_i)	Frequency(f_i)	$f_i x_i$
3-5	4	5	20
5-7	6	10	60
7-9	8	10	80
9-11	10	7	70
11-13	12	8	96
Total		40	326

$$\therefore \text{Mean} = \frac{\sum f_i x_i}{\sum f_i} = \frac{326}{40} = 8.15 \text{ litres}$$

iii. (b) 16300 litres

Solution:

If 2000 vehicles comes daily and average quantity of diesel required for a vehicle is 8.15 liters, then total quantity of diesel required,
 $= 2000 \times 8.15 = 16300 \text{ liters}$

iv. (c) 16

Solution:

Here, $N = 40$ and $\frac{N}{2} = 20$

c.f. for the distribution are 5, 15, 25, 32, 40

Now, cf just greater than 20 is 25 which is corresponding to the class interval 7 - 9.

So median class is 7 - 9.

$\therefore$ Required sum of upper limit and lower limit $= 7 + 9 = 16$

v. (b) 7.7 litres

Solution:

We know, Mode $= 3 \text{ Median} - 2 \text{ Mean}$

$$= 3(8) - 2(8.15) = 24 - 16.3 = 7.7$$

2. Answer:

i. (d) $2m - b$

Solution:

We know that,

$$\text{Class mark} = \frac{\text{Lower limit} + \text{Upper limit}}{2}$$

$$\Rightarrow m = \frac{\text{Lower limit} + b}{2} \Rightarrow \text{Lower limit} = 2m - b$$

ii. (a) 341hrs

Solution:

Lifetime (in hours)	Class mark (x_i)	(f_i)	$d_i = x_i - A$	$f_i d_i$
150-200	175	14	-150	-2100
200-250	225	56	-100	-5600
250-300	275	60	-50	-3000
300-350	325 = A	86	0	0
350-400	375	74	50	3700
400-450	425	62	100	6200
450-500	475	48	150	7200
Total		50		6400

$\therefore$ Average lifetime of a packet

$$= A + \frac{\sum f_i d_i}{\sum f_i} = 325 + \frac{6400}{400} = 341\text{hrs}$$

iii. (b) 340hrs

Solution:

$$\text{Here, } N = 400 \Rightarrow \frac{N}{2} = 200$$

Also, cumulative frequency for the given distribution are 14, 70, 130, 216, 290, 352, 400

$\therefore$ c.f just greater than 200 is 216, which is corresponding to the interval 300-350.

$$l = 300, f = 86, \text{c.f.} = 130, h = 50$$

$$\therefore \text{median} = l + \left(\frac{\frac{N}{2} - \text{c.f.}}{f} \right) \times h = 300 + \left(\frac{200 - 130}{86} \right) \times 50$$

$$= 300 + 40.697 \approx 340.697 \approx 340\text{hrs (approx.)}$$

iv. (a) 340hrs

Solution:

We know that Mode = 3 Median - 2 Mean

$$= 3(340.697) - 2(341)$$

$$= 1022.091 - 682 = 340.091 \approx 340\text{hrs}$$

v. (c) 340hrs

Solution:

Since, minimum of mean, median and mode is approximately 340hrs. So, manufacturer should claim that lifetime of a packet is 340hrs.

Assertion Reason Answer-

(c) A is true but R is false.

(c) A is true but R is false.