ST. LAWRENCE HIGH SCHOOL

A JESUIT CHRISTIAN MINORITY INSTITUTION

CLASS 8

SUBJECT : Arithmetic

Study Material 5

Data Handling

Date:8.5.2020

DATA HANDLING

KEY FACTS

1. A collection of numerical facts about objects or events is called **data.**
2. **Statistics** is the science of collecting, organising and analyzing sets of data in order to reveal information.
3. Arranging raw data in ascending or descending order of magnitude is called an **array.**
4. The difference between the greatest and the least values of observation in a set of data is called the **range** of that set.
5. The number of times an observation occurs in a given data set is called its **frequency.**
6. **Grouped frequency distribution:** If the data extends over a wide range, data is usually condensed into groups or classes called the **class intervals.**

   (i) Each class interval is bounded by two figures which are called **class limits.**

   Ex. 0 – 5, 5 – 10, 10 – 15 are class intervals with class limits as 0, 5, 10, 15 etc.

   For a particular class interval, the number on the left hand side is the lower limit and the number on the right hand side is the upper limit.

   Ex. In the interval 10 –15, 10 is the lower limit and 15 is the upper limit.

   There are two types of class intervals:

   (a) **Exclusive form:** Here the upper limit of one class coincides with the lower limit of the next class.

   Ex. 0 – 10, 10 – 20, 20 – 30, ... etc.

   Note. 5–10 means 5 and less than 10. 10 is included in the class interval 10 –15.

   (b) **Inclusive form:** Here the upper limit of a class interval does not coincide with the lower limit of the next class.

   Ex. 1– 10, 11 – 20, 21 – 30, ... etc.

   To convert the class interval from an inclusive form to exclusive form, we find the difference between the upper limit of a class interval and the lower limit of the next class interval and divide it by 2.

   Ex. In the above example, diff $= \frac{11-10}{2} = \frac{1}{2} = 0.5$.

   Thus, the class limits are converted to exclusive form by subtracting 0.5 from the lower limit and adding 0.5 to the upper limit.

   Hence, the class boundaries are 0.5 – 10.5, 10.5 – 20.5, ...

   (ii) The difference between the upper and lower class limits is called the **size** or **width** of the **class interval.**

   Ex. Size of the class interval 0–10 = 10 – 0 = 10.

   (iii) The mid-value of a class-interval is called its **class mark.**

   Class mark = upper limit + lower limit

   Ex. Class mark of the interval 10 –15 $= \frac{10+15}{2} = 12.5$

(iv) The frequency corresponding to a particular class interval is called its **class frequency.**

7. Arithmetic mean

(i) Arithmetic mean for n observations $x_1, x_2, \ldots x_n$ is given by

$$\frac{x_1 + x_2 + \ldots + x_n}{n}$$

(ii) Arithmetic mean for a discrete series $= \dfrac{\Sigma(f_i x_i)}{\Sigma f_i}$, where x_i is the observation and f_i is the corresponding frequency.

(iii) Arithmetic mean of grouped data $= \dfrac{\Sigma f_i x_i}{\Sigma f_i}$, where x_i is the class mark of interval corresponding to the class frequency f_i.

(iv) If a particular number is multiplied, divided, added or subtracted to each observation, then the mean also gets multiplied, divided, added or subtracted to that number.

8. Median of ungrouped data:

To find the median of an individual series, arrange the given numbers in ascending or descending order. If the number of observations is:

(a) odd, the median is size of the $\left(\dfrac{n+1}{2}\right)$th term.

(b) even, the median is size of $\left[\dfrac{\left(\dfrac{n}{2}\right)\text{th term} + \left(\dfrac{n+1}{2}\right)\text{th term}}{2}\right]$.

9. **Mode:** The most occurring observation of a given set of data is called its mode. The mode is the term that has the largest frequency.

10. **Empirical formula:** Mode = 3 Median – 2 Mean

INTRODUCTION

The word '**statistics**' is derived from the Latin word '**status**', which means **political state**. Political states had to collect information about their citizens to facilitate governance and plan for development. Then, in course of time, statistics came to mean a branch of mathematics which deals with the collection, classification, and analysis of numerical data.

In this chapter, we shall learn about the classification of data viz., grouped and ungrouped, measures of central tendency, and their properties.

Data

The word data means information in the form of numerical figures or a set of given facts. For example, the percentage of marks scored by 10 pupils of a class in a test is:

72, 84, 82, 96, 94, 98, 99, 67, 92 and 93.

The set of these figures is the data related to the marks obtained by 10 pupils in a class test.

Types of Data

Statistics is basically the study of numerical data. It includes methods of collection, classification, presentation, analysis of data, and inferences from data. Data as such can be qualitative or quantitative in nature. If one speaks of honesty, beauty, colour, etc., then the data is qualitative, whereas height, weight, distance, marks, etc., are quantitative. Data can also be classified as raw data and grouped data.

Raw Data Data obtained from direct observation is called raw data.

The marks obtained by 10 students in a monthly test are an example of raw data or ungrouped data.

Note The difference between the highest and the lowest values of the data is called 'Range'.

E.g., Range of the data of the first 25 natural numbers is $25 - 1 = 24$.

In fact, very few can be inferred from this data. So, to make this data clearer and more meaningful, we group it into ordered intervals.

Grouped Data To present the data in a more meaningful way, we condense the data into convenient number of classes or groups, generally not exceeding 10 and not less than 5. This helps us in perceiving, at a glance, certain salient features of data.

Some Basic Definitions

Before getting into the details of tabular representation of data, let us review some basic definitions.

1. **Observation:** Each numerical figure in a data is called an observation.
2. **Frequency:** The number of times a particular observation occurs is called its frequency.

Tabulation or Presentation of Data

When data is arranged systematically in a tabular form, it is called tabulation or presentation of the data. These grouping results in a table called the frequency table, which indicates the number of scores within each group. Many conclusions about the characteristics of the data, the behaviour of variables, etc., can be drawn from this table.

The quantitative data that is to be analysed statistically can be divided into three categories:

1. Individual series
2. Discrete series
3. Continuous series

Individual Series

Any raw data that is collected forms an individual series.

Examples: (a) The heights of 6 students:
120, 126, 132, 136 138, and 142 (in cm)

(b) Percentage of marks obtained by 10 students in a test:
78, 42, 93, 47, 82, 67, 92, 79, 94, and 93

Discrete Series

A discrete series is formulated from raw data by taking the frequency of the observations into consideration.

Example: Given below is the data showing the number of computers in 12 families of a locality.

1, 1, 2, 3, 2, 1, 4, 3, 2, 2, 1, 1

Arranging the data in the ascending order.

1, 1, 1, 1, 1, 2, 2, 2, 2, 3, 3, 4

To count, we can use tally marks. We record tally marks in bunches of five, the fifth one crossing the other four diagonally, i.e., 𝍸 .

Thus, we may prepare a frequency table as below:

Number of Computers	Tally Marks	Number of Families (Frequency)
1	𝍸	5
2	IIII	4
3	II	2
4	I	1

Continuous Series

When the data contains large number of observations, we put them into different groups called class intervals such as 1–10, 11–20, 21–30, etc.

Here, 1–10 means data whose values lie between 1 and 10 including both 1 and 10.

This form is known as inclusive form. Also, 1 is called the **lower limit** and 10 is called the **upper limit**.

Example: Given below are the marks (out of 50) obtained by 30 students in an examination:

43	19	25	32	48
17	29	9	15	50
7	24	20	37	44
22	2	50	27	25
18	42	16	1	33
25	35	45	35	28

Taking class intervals 1–10, 11–20, 21–30, 31–40, and 41–50, we construct a frequency distribution table for the above data.

First, we write the marks in the ascending order as:

1	2	7	9	15	16	17	18	19	20
22	24	25	25	25	27	28	29	32	33
35	35	37	42	43	44	45	48	50	50

Now, we can prepare the frequency distribution table as below:

Class Interval	Tally Marks	Frequency
1–10	IIII	4
11–20	𝍸 I	6
21–30	𝍸 III	8
31–40	𝍸	5
41–50	𝍸 II	7

Now, with this idea about tabulation, let us review some more concepts.

Class Interval

A group into which the raw data is condensed is called a class-interval.

Each class is bounded by two figures, which are called the class limits. The figure on the LHS is called the lower limit and the figure on the RHS is called the upper limit of the class. Thus, 0–10 is a class with lower limit being 0 and upper limit being 10.

Class Boundaries

In an exclusive form, the lower and upper limits are known as class boundaries or true lower limit and true upper limit of the class, respectively. In order to convert the class interval from inclusive form to exclusive form, firstly we have to find the difference between the upper limit of class interval and the lower limit of the next class interval and divide it by two. Let it be denoted by a. Thus, in the following example, $a = (11 - 10)/2 = 0.5$. The boundaries in an exclusive form are obtained by subtracting 0.5 from the lower limit and adding 0.5 to the upper limit of the inclusive form.

Inclusive Form	Exclusive Form
1–10	0.5–10.5
11–20	10.5–20.5
21–30	20.5–30.5
31–40	30.5–40.5

Class Size

The difference between the true upper limit and the true lower limit is called the class size. Hence, the class size 11–20 (Inclusive form) = 20.5–10.5 = 10.

Class Mark or Mid-Value

$$\text{Class mark} = \frac{1}{2}\ (\text{upper limit} + \text{lower limit})$$

Thus, the class mark of 11–20 is $\frac{1}{2}(20 + 11) = 15.5$.

SOLVED EXAMPLES

Given below is the data showing the number of children in a family in a locality of 15 families. Prepare the frequency table for the data 1, 3, 5, 4, 3, 2, 1, 1, 3, 2, 2, 1, 4, 2, and 1.

SOLUTION

Arranging the data in ascending order, we have:

1, 1, 1, 1, 1, 2, 2, 2, 2, 3, 3, 3, 4, 4, 5

To count, we can use tally marks. We record tally marks in bunches of five, the fifth one crossing the other diagonally, i.e., 𝍸.

Hence, we may prepare a frequency table as shown below.

Number of Children	Tally Marks	Number of Families (frequency)
1	𝍸	5
2	IIII	4
3	III	3
4	II	2
5	I	1

Find the mean of the first 10 prime numbers.

SOLUTION

The first 10 prime numbers are 2, 3, 5, 7, 11, 13, 17, 19, 23, 29.

$$\text{Arithmetic mean (A. M)} = \frac{\text{Sum of observations}}{\text{Total number of observations}}$$

$$= \frac{2+3+5+7+11+13+17+19+23+29}{10} = \frac{129}{10} = 12.9$$

Find the median of the data 21, 33, 15, 19, 26, 32, 14, and 24.

SOLUTION

Arranging the given values in the ascending order, we have 14, 15, 19, 21, 24, 26, 32, 33.

As the given number of observations is even ($n = 8$), then the average of the two middle terms is the median, i.e., $\left(\frac{n}{2}\right)$th and $\left(\frac{n}{2}+1\right)$th observations average.

i.e., median is the average of the 4th and the 5th terms.

$$\therefore \text{Median} = \frac{21+24}{2} = 22.5$$

Find the median of the following observations 12, 18, 11, 21, 32, 16, and 22.

SOLUTION

Arranging the given observations in the ascending order, we have 11, 12, 16, 18, 21, 22, 32.

The middle term of the above data is $\left(\frac{n+1}{2}\right)$th observation (here $n = 7$).

i.e., 4th observation which is 18.

$\therefore$ Median = 18

The median of 11 observations is 10. Find the maximum number of possible observations in the data which are less than 10.

SOLUTION

Median of 11 observations is 10. $\Rightarrow$ Middle most value is 10.

$\Rightarrow$There are 5 observations before 10 and after 10.

Mode

The most frequently found value in the data is called the mode. This is the measure which can be identified in the simplest way.

The mode of the unimodal data 7, 8, 9, 8, 9, 10, 9, 10, 11, 10, 11, 12, and x is 10. Find the value of x.

SOLUTION

The given observations are 7, 8, 9, 8, 9, 10, 9, 10, 11, 10, 11, 12, and x.

Observations	Frequency
7	1
8	2
9	3
10	3
11	2
12	1
x	1

Also given that the mode is 10 and the given data is unimodal.

$\therefore x = 10$

If the mode of the following data is 4, then find the median of the data.

x	2	3	P	5	6
f	1	2	7	6	6

HINTS

(a) Use the definition of mode and proceed.

(b) As p has highest frequency, mode = p, i.e., $p = 4$.

(c) There are a total of 22 observations, so the median is the average of 11th and 12th observations.

Notes

1. For a given data, the mode may or may not exist. In a series of observations, if no item occurs more than once, then the mode is said to be ill-defined.
2. If the mode exists for a given data, it may or may not be unique.
3. Data having a unique mode is unimodal, whereas data having two modes is bi-modal.

The daily wages of 20 workers are given below:

Salary (in rupees)	Number of workers
60	4
80	2
100	2
150	8
200	4
Total	20

Find the mean daily wage of twenty workers.

SOLUTION

The mean $\bar{x}$ is given by $\bar{x} = \dfrac{(60 \times 4) + (80 \times 2) + (100 \times 2) + (150 \times 8) + (200 \times 4)}{20}$

$= ₹130$

i.e., the mean salary of the workers is ₹130.

Find the mode from the following data:

Variable (x)	5	10	15	20	25	30	35
Frequency (f)	3	4	10	7	6	4	2

SOLUTION

From the given data, the variable with highest frequency (10) is 15.

$\therefore$ Mode is 15.

If the mean of the data 9, 17, 18, 14, x, 16, 15, 11, and 12 is x, then find the value of x.

SOLUTION

Given, mean $= x$

$$\frac{9+17+18+14+x+16+15+11+12}{9} = x$$

$$\frac{112+x}{9} = x$$

$$9x = 112 + x$$

$$8x = 112$$

$$x = \frac{112}{8}$$

$\therefore x = 14$

The mean of y and $1/y$ is z. Find the mean of y^2 and $\frac{1}{y^2}$.

SOLUTION

Mean of y and $\frac{1}{y}$ is $\frac{y+\frac{1}{y}}{2} = z$

$\Rightarrow y+\frac{1}{y} = 2z$

Squaring on both sides, we get

$$y^2+\frac{1}{y^2}+2 = 4z^2$$

$$y^2+\frac{1}{y^2} = 4z^2-2$$

$$\frac{y^2+\frac{1}{y^2}}{2} = 2z^2-1$$

$\therefore$ The required mean $= \frac{y^2+\frac{1}{y^2}}{2} = 2z^2-1$

Find the median of the following data:

Variable (x)	2	3	5	7	11	13	17
Frequency (f)	4	**2**	6	9	10	3	2

SOLUTION

Variable	Frequency	Cumulative Frequency
2	4	4
3	2	4 + 2 = 6
5	6	4 + 2 + 6 = 12
7	9	4 + 2 + 6 + 9 = 21
11	10	4 + 2 + 6 + 9 + 10 = 31
13	3	4 + 2 + 6 + 9 + 10 + 3 = 34
17	2	4 + 2 + 6 + 9 + 10 + 3 + 2 = 36

Here, $N = 36$

$\therefore$ Median $= \left(\frac{36}{2}\right)$th term = 18th term = 7

QUESTION BANK

TEST YOUR CONCEPTS

Very Short Answer Type Questions

1. The class size of 9–12 (Inclusive form) is _____.
2. If the mean of x, y, and z is k, then mean of $\frac{x}{y}$, 1 and $\frac{z}{y}$ is _____.
3. A bar graph is drawn to the scale 1 cm = 250 units. The length of the bar representing a quantity of 300 units is _____.
4. In bar graphs, the width of the bar represents the value of the item. (True/False)
5. If the median of a, b, c, d, e, and f ($a < b < c < d < e < f$) is k, then the median of b, c, d, and e is _____.
6. In histograms, width of the bars is determined by class size. (True/False)
7. The class mark of 10–15 (exclusive form) is _____.
8. The mode of 1, 2, 3, 4, 6, 6, 6 7, 8, 9, and x is _____.
9. If the mean of p, q, r, s, and t is 5, then $(p-5) + (q-5) + (r-5) + (s-5) + (t-5) =$ _____.
10. The number of times a particular observation occurs is called its _____.
11. Which measure of central tendency can be used to know the sales strategy in a shoe shop?
12. The mean of the first n natural numbers is 10, then the number of observations is _____.
13. If the total weight of 25 students is 625 kg, then mean weight of 25 students is _____.
14. In a certain data, highest value is $(x + 4)$ and lowest value is $(x - 4)$. Range of the data is _____.
15. The mean of certain number of observations is 10. Even though one observation is deleted, mean is not altered. Then the deleted observation is _____.
16. For a certain data, if the median is equal to the mean, then the mode = _____.
17. A person spends 20% of his income towards food, 25% for education, 15% for miscellaneous expenses, and saves 40%. The angle of sector representing expenditure towards education in pie diagram showing his income is _____.
18. The frequency of an observation p is q and the frequency of another observation q is p, then mean is _____.
19. The most stable measure of central tendency is _____.
20. Krishna wrote 1 to 100 numbers, the frequency of the digit 9 is _____.
21. Find the mean of the following data.

    4, 6, 7, 13, 25, and 35
22. Find the median of the following data.

    24, 14, 13, 26, 12, 18, and 9
23. If Mean = 24 and Median = 26, then find the Mode.
24. Find the mode of the following data.

    11, 12, 11, 12, 13, 11, 12, 13, and 14.
25. Find the mean of the following data.

    23, 24, 27, 31, 37, 46, and 78.
26. Find the median of the following data.

    72, 93, 48, 36, 79, 23, 41, and 81.
27. Find the mode of the following data.

    2, 4, 2, 3, 3, 4, 6, 8, 6, 2, 2, and 3

28. Find the class marks and class size of the following distribution.

Class Interval	0–10	10–20	20–30	30–40	40–50

29. Percentage of marks of 35 students is given in the following grouped frequency distribution table.

Marks (in %)	40–50	50–60	60–70	70–80	80–90	90–100
Number of students	5	4	7	**8**	**6**	**5**

Which class has highest frequency?

30. If Median = 29 and Mode = 31, then find the Mean.

ANSWERS

TEST YOUR CONCEPTS

Very Short Answer Type Questions

1. 4
2. $\frac{k}{y}$
3. 1.2 cm
4. False
5. k
6. True
7. 12.5
8. 6
9. 0
10. frequency
11. Mode
12. 19
13. 25 kg
14. 8
15. 10
16. mean
17. 90^o
18. $\frac{2pq}{p+q}$
19. median
20. 20
21. 15
22. 14
23. 30
24. 11 and 12
25. 38
26. 60
27. 2
28. 10
29. 70 – 80

SELF ASSESSMENT EXERCISE

MATHEMATICAL REASONING

DIRECTION (1-2) : Study the following table which shows the production of three different types of cars over the years.

Years	Production of Car P	Production of Car Q	Production of Car R
2001	76	59	28
2002	82	62	36
2003	65	47	42
2004	70	54	31
2005	85	57	49
2006	80	68	38

1. The average production of which of the following types of cars was maximum?

(A) Q (B) P
(C) R (D) All are equal

2. The difference between the total production of three cars in the year 2004 and 2006 is _____.

(A) 11 (B) 43
(C) 31 (D) 28

3. The heights of 10 students were measured in cm and the results are as follows : 147, 139, 135, 136, 149, 166, 152, 163, 155, 144.

(i) What is the range of the data?
(ii) What is the mean height of the students?
(iii) How many students are there whose height is more than the mean height?

	(i)	(ii)	(iii)
(A)	30	100.5	3
(B)	31	148.6	5
(C)	32	149	5
(D)	28	148.6	4

4. The mode and median of the following data 21, 23, 25, 23, 21, 24, 22, 20, 23 respectively, are

(A) 23, 23 (B) 21, 23
(C) 25, 21 (D) 23, 21

5. Mean of 9 observations was found to be 35. Later on, it was detected that an observation 81 was misread as 18, then the correct mean of the observations is _____.

(A) 40 (B) 41
(C) 42 (D) 43

6. If the mean of 6, 8, 5, x and 4 is 7, then the value of x is _____.

(A) 11 (B) 12
(C) 13 (D) 14

7. The mode of the following distribution is _____.

Size	2	3	4	5	6	7	8
Frequency	10	12	25	20	25	15	11

(A) 2 (B) 8
(C) Both 4 and 6 (D) 5

SOLUTIONS

1. **(B)** : Average production of Car P

$$= \frac{\text{Sum of production of Car P in each year}}{\text{Total number of years}}$$

$$= \frac{76+82+65+70+85+80}{6} = \frac{458}{6} = 76.33$$

Average production of Car Q

$$= \frac{59+62+47+54+57+68}{6} = \frac{347}{6} = 57.83$$

Average production of Car R

$$= \frac{28+36+42+31+49+38}{6} = \frac{224}{6} = 37.33$$

Clearly, average production of Car P is greater than Car Q and Car R.

Hence, average production of Car P is maximum.

2. **(C)** : Total production of cars in year 2004 = 70 + 54 + 31 = 155

Total production of cars in year 2006 = 80 + 68 + 38 = 186

Hence, difference between the total production of three cars in the year 2004 and 2006 = 186 – 155 = 31

3. **(B)** : (i) Range = Maximum value – Minimum value = 166 – 135 = 31

(ii) Mean height of the students

$$= \frac{\text{Sum of height of all the students}}{\text{Total number of students}}$$

$$= \frac{147+139+135+136+149+166+152+163+155+144}{10}$$

$$= \frac{1486}{10} = 148.6$$

(iii) Height of students more than the mean height are 149, 166, 152, 163 and 155.

i.e., 5 student's height is more than the mean height.

4. **(A)** : Here 23 occurs most frequently.

∴ Mode = 23

On arranging terms in ascending order, we get 20, 21, 21, 22, 23, 23, 23, 24, 25

Total number of terms = 9

$$\therefore \text{ Median } = \left(\frac{9+1}{2}\right)^{th} \text{ term} = 5^{th} \text{ term} = 23$$

5. **(C)** : Incorrect mean of 9 observations = 35

Incorrect sum of 9 observations = 35 × 9 = 315

Since, 81 is misread as 18.

So, correct sum of observations = 315 – 18 + 81 = 378

$$\therefore \text{ Correct mean } = \frac{378}{9} = 42$$

6. **(B)** : Mean = 7, *i.e.*, $\frac{6+8+5+x+4}{5} = 7$

$$\Rightarrow \quad x + 23 = 35 \Rightarrow \quad x = 12$$

7. **(C)** : Mode is that observation which have highest frequency. Since, both 4 and 6 have highest frequency.

∴ Option (C) is correct.

INDRANIL GHOSH