# Data Handling

## Introduction

Modern society is information oriented. Every person wants numeric information of different fields of the society like the marks obtained in a particular subject by the students, five year plans etc.
Statistics is a branch of Mathematics which deals with the process, analyzing and interpreting the data. Data is defined as the particular information in numeric form.

### Primary Data

Primary data means the data that have been collected by collector for some purpose.
It means when an authorized organization or an investigator or an enumerator collects the data for the use to himself or with the help or an institution or an expert then the data collected is called primary data.

### Secondary Data

Secondary data is data that have been collected by others and used by someone else. It means that after performing statistical operations of primary data the result become useful to others is called the secondary data.

### Raw Data

Raw data (or ungrouped data) is the data which is not arranged in any particular fashion or pattern. It is the original form of the data.
e.g. The height of 5 students in a class is 123 cm, 120 cm, 129 cm, 135 cm, 121 cm.

### Grouped Data

Grouped data is the data which is arranged in classes or group to bring out salient mature of the group. Firstly it may arrange in ascending or descending order and then divide into groups.

### Variable

A measurable characteristic is called a variable or variate  eg. Age, height, income

### Attribute

A non - measurable characteristic is called an attribute.

### Frequency

The number of times a particular observation occurs in a data is called frequency. In other words, it is number of times an observation occurs, e.g in a data 1, 2, 3,4,5, 2, 1, 3, 2, 4, 6, 5, 2, 3, 2, 1, 2, 5.
The frequency of 2 in the above data is 6 because it occurs 6 times in the observation.

## Illustrative EXAMPLE

The run scored by 11 member of a cricket team are: 34, 0, 25, 34, 67, 73, 67,1, 0,71
Represent the given data by using tally marks.

**Solution:**

| Scores | Tally mark | Frequency |
|---|---|---|
| 0 | II | 2 |
| 1 | I | 1 |
| 25 | I | 1 |
| 34 | III | 3 |
| 67 | II | 2 |

| | | |
|---|---|---|
| 71 | I | 1 |
| 73 | I | 1 |

## Mean

Mean is defined as the ratio of sum of observation to total number of observation.

$$mean = \frac{sum\,of\ observation}{Totel\ number\ of\ observation} \text{ or , } \overline{X} = \frac{\sum_{i=1}^{n} x_i}{n}$$

**Illustrative EXAMPLE**

Find the mean of first five even numbers

**Solution:**

First five even numbers are 2, 4, 6, 8, 10

$$mean = \frac{2+4+6+8+10}{5} = 6$$

## Median

**Step 1:** Arrange the given data either in ascending or descending order

**Step 2:** Count the number of observation.

**Step 3:** If the number of observation is odd, then

$$median = \left(\frac{n+1}{2}\right)^{th} observation$$

**Step 4:** If the number of observation is even, then

$$median = \frac{\left(\frac{n}{2}\right)^{th} observation + \left(\frac{n}{2}+1\right)^{th} observation}{2}$$

**Illustrative EXAMPLE**

Find the median of 4, 5, 3, 3, 2, 1, 5

**Solution:**

Arrange the data in ascending order, we get

1, 2, 3, 3, 4, 5, 5 Here the number of observation is 7 which is odd number that is Median $= \left(\frac{n+1}{2}\right)^{th}$ observation

$= \left(\frac{7+1}{2}\right)^{th}$ observation i.e the fourth observation. Therefore, median is 3

**Example**

Find the median of first six prime number

**Solution:**

First six prime numbers are 2, 3, 5, 7, 11, 13

$$median = \frac{\left(\frac{6}{2}\right)^{th} observation + \left(\frac{6}{2}+1\right)^{th} observation}{2}$$

$$= \frac{third\,observation + fourth\,obsercation}{2} = \frac{5+7}{2} = 6$$

## Mode

It is defined as the value of variable which has highest frequency.
The mode of data 10, 12, 14, 10, 14, 20, 30, 10, 25, 23, 10, 12 is 10, because it occurs most of the time.
In the class - interval 1- 5, 5 is called upper limit and 1 is the lower limit.

## Commonly Asked QUESTIONS

---

**David throws a die 10 times and the following are outcomes 2, 5, 6, 6,1, 5, 4,1, 4, 6, 5. The mean of the above observation is:**

(a) 3 (b) 4.5
(c) 5 (d) 45
(e) None of these

**Answer:** (b)

**Explanation**

$$mean = \frac{sum\,of\ observation}{Total\ number\ of\ observation}$$

$$= \frac{2+5+6+6+1+5+4+1+4+6+5}{10} = 4.5$$

---

**If a, b, c, d, and e are five consecutive odd numbers, then their mean is:**

(a) b (b) c
(c) e (d) a
(e) None of these

**Answer:** (b)

**Explanation**

Let the consecutive odd numbers a, b, c, d and e are as follows

$a = 2x+1, b = 2x+3, c = 2x+5,\ d = 2x+7$ and $e = 2x+9$

$$mean = \frac{a+b+c+d+e}{5} =$$

$$\frac{(2x+1)+(2x+3)+(2x+5)+(2x+7)+(2x+9)}{5} = c$$

---

**The median of first five consecutive even numbers p, q, r, s and t is:**

(a) $(q+s)\div 2$ (b) p
(c) $(p+q)\div 2$ (d) $(p+q+r)\div 3$
(e) None of these

**Answer:** (a)

---

**The mean, median and mode of the following data are respectively: 5, 17, 21, 21, 7,13,1, 3**

(a) 10, 10, 21 (b) 11, 21, 2
(c) 11, 10, 21 (d) 11, 10, 5
(e) None of these

**Answer:** (c)

---

**Stuart performs his project work on the topic that the number of students likes soft drink of different flavours in a school. After collecting the data he wants to know the most flavoured soft drink which is liked by most of the students. Which central tendency makes his wish true?**

(a) Mean (b) Row data
(c) Median (d) Mode
(e) None of these

**Answer:** (d)

## Mean of Grouped Data

$$\text{Mean} = \frac{\sum_{i=1}^{n} x_i f_i}{\sum_{i=1}^{n} f_i} \text{ where}$$

$$= \sum_{i=1}^{n} x_i f_i = x_i f_i + x_2 f_2 + x_3 f_3 + x_4 f_4 + \ldots\ldots\ldots$$

And $\sum_{i=1}^{n} f_i = f_1 + f_2 + f_3 + f_4 + \ldots\ldots$

The difference between upper and lower limit of a class interval is called class size.

**Range**

The difference between maximum and minimum value of the observation is called range.

**Class Interval**

A data can be classified into different intervals for convenience to analyzing it. The interval in which variates lies is called class interval.

**Class Mark**

C.M. $= \frac{1}{2}$ (lower limit + upper limit)

## Bar Graph

It is the most simple and widely used chart (or graph).
In this representation, bars are drawn with some width at each column. The scale showing the heights of the bars is represented along with the axis and in proportion to the frequency.

## Histogram

It.is a graphical representation of frequency distribution in which class-intervals are taken along x - axis and frequency are taken along y - axis.

## Frequency Polygon

The curve obtained by joining midpoints of successive tops of the rectangles of a histogram represents frequency polygon.

## SUMMARY

- Data is defined as the particular information in numeric form.
- Mean $= \dfrac{sum\,of\ observation}{Total\ number\ of\ observation}$
- If the number of observations is odd then median $= \left(\dfrac{n+1}{2}\right)^{th}$ observation
- If the number of observations is even

  Median $= \dfrac{\left(\dfrac{n}{2}\right)^{th} observation + \left(\dfrac{n}{2}+1\right)^{th} observation}{2}$
- Mode is defined as the value of variable which has highest frequency.

## You Must KNOW

- Stat-etymology was the early applications of statistical thinking revolved around the needs of states to base policy on demographic and economic data.
- Statistics help in prediction and forecasting using data and statistical models
- There are four main levels of measurement used in statistics: ratio, interval, nominal and ordinal.

# Self Evaluation TEST

Duration 10 Minutes

---

**1.** **The mean age of 30 students of a class is 10 years. If the age of their teacher is also included then the mean is increased by one year. What is the age of the teacher?**

(a) 21 years (b) 22 years
(c) 41 years (d) 42 years
(e) None of these

---

**2.** **The traffic police recorded the speed (in km/h) of 10 motors as 47, 53, 49, 60, 39, 42, 57, 55, 48 and 52. Later on an error in the recording instrument was found. Find the correct mean speed of the motors if the instrument recorded the speed 6 km/h less is each case.**

(a) 55.2km/h (b) 56.2 km/h
(c) 57.2 km/h (d) 58.2 km/h
(e) None of these

---

**3.** **The mean temperature of Monday, Tuesday and Wednesday is $40°C$. The mean temperature of Tuesday, Wednesday and Thursday is $41°C$. If the temperature of Thursday is $30°C$ ,then the temperature on Monday is:**

(a) $25°C$ (b) $30°C$
(c) $35°C$ (d) $27°C$
(e) None of these

---

**4.** **A student scored the following marks in an examination in Hindi 75, Mathematics 60, English 59 and Engg. Drawing 63. Find the weighted mean if the approved weightage of students in the above subjects are 2, 4, 1 and 3 respectively.**

(a) 61.8 marks (b) 60.1 marks
(c) 62.8 marks (d) 63.8 marks
(e) None of these

---

**5.** **Find the median of 11, 15, 13, 27, 19, 24 and 20. If 13 is replaced by 31 then find the new median.**

(a) 10 (b) 20
(c) 30 (d) 40
(e) None of these

---

**6.** **The median of the following observations arranged in ascending order is 24 find $x$.**

$11,12,14,18,x+2,x+4,30,32,35$ **and** $41$.

(a) 11 (b) 20
(c) 21 (d) 22
(e) None of these

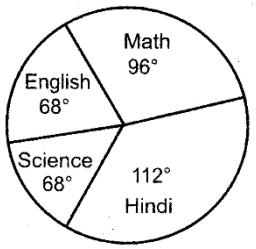
**7.** **If the following data is represented in a pie diagram then what will be the angle of the sector corresponding to boys?**

| Males | Females | Girls | Boys | Total |
|---|---|---|---|---|
| 2500 | 2000 | 1500 | 4000 | 10,000 |

(a) 120°
(b) 90°
(c) 144°
(d) 25°
(e) None of these

---

**8.** **Find out the marks in Hindi from pie chart given below, if the total marks be 540.**




(a) 168
(b) 126
(c) 102
(d) 144
(e) None of these

---

**9.** **The diagrammatic representation with the help of pictures is called:**

(a) Histogram
(b) Pie chart
(c) Pictogram
(d) Bar chart
(e) None of these

---

**10.** **Find the mean of following distribution.**

| X | 10 | 15 | 20 | 25 | 30 | 35 | 40 |
|---|---|---|---|---|---|---|---|
| f | 7 | 7 | 13 | 15 | 7 | 4 | 5 |

(a) 23.4412
(b) 23.34
(c) 23.45
(d) 23.55
(e) None of these

---

**Answers – Self Evaluation Test**

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **1.** | C | **2.** | B | **3.** | D | **4.** | D | **5.** | B | **6.** | C | **7.** | C | **8.** | A | **9.** | C | **10.** | C |

# Self Evaluation Test

## SOLUTIONS

---

**1.** Total age of 30 students $= 10 \times 30 = 300$ years
Mean age of 30 students and the teacher = 11 years
i.e. total age of 30 students and the teacher $31 \times 11 = 341$ years.
Hence, age of the teacher $= (341 - 300) = 41$ year

---

**2.** Mean speed using error instrument

$$\frac{47+53+49+60+39+42+57+55+48+52}{10} = \frac{502}{10} = 50.2\,km/m$$

Corrected mean speed

$$\frac{53+59+55+66+45+48+63+61+54+58}{10} = \frac{562}{10} = 56.2\,kh/\text{h}$$

---

**3.** Total temperature of Monday, Tuesday and Wednesday $= 40 \times 3 = 120°C$
Total temperature of Tuesday, Wednesday and Thursday $= 41 \times 3 = 123^{o}C$
Temperature of Thursday $= 30^{o}\ C$, on adding (i) and (iii) and subtracting
(ii) use get the temperature of Monday $= (120 + 30 - 123)^{o}C = 27^{o}C$

---

**4.** Weighted mean $\frac{75 \times 2 + 60 \times 4 + 59 \times 1 + 63 \times 3}{2+4+1+3}$

$$\frac{150+240+59+189}{10} = \frac{638}{10} = 63.8 \text{ marks}$$

---

**5.** Arranging the given data in ascending order, 11,15, 13, 27,19, 24, 20, here no. of observation = 7 (odd)
$\therefore$ median = 19, when 13 is replaced by 31, the data becomes as follows on arranging in ascending order.
11, 15, 19, 20, 24,27, 31
$\therefore$ median = value of $4^{th}$ observation
Since $4^{th}$ observation in new series is 20, new median is 20.

---

**6.** No. of observation n = 10 term,
Value of $\left(\frac{10}{2}\right)th$ observation + value of $\left(\frac{10}{2}+1\right)th$ the observation

$$\therefore \text{ Median } \frac{Value\,of\left(\frac{10}{2}\right)th\,observation + Value\,of\left(\frac{10}{2}+1\right)th\ observation}{2}$$

---

**7.** Required angle $= \frac{4000}{10,000} \times 360^o = 144^o$

**8.** Marks obtained in Hindi $= \frac{112}{360} \times 540 = 168$

**9.** Pictogram is an ideogram that conveys its meaning through its pictorial reasonableness to a physical object.

**10.** Mean $= \frac{\sum x_i f_i}{\sum f_i} = \frac{70+105+260+375+210+140+200}{58} = 23.45$