

# Statics

## Some Terms related to Statistic

1. **Primary data:**

The information collected by the investigator himself or herself with a definite purpose in his or her mind is called primary data.

2. **Secondary data:**

The information gathered from a source which already had the information stored is called secondary data.

3. **Raw data:**

The numerical data recorded in its original form as it is collected by the investigator or received from some source is called raw data.

4. **Variable:**

A quantity which is being measured in an experiment (or survey) is called a variable. Height, age and weight of people, income and expenditure of people, number of members in a family, number of workers in a factory, marks obtained by students in a test, the number of runs scored in a cricket match etc., are examples of variables.

Variables are of two types:

a. **Continuous variable:** A variable which can take any value between two given values is called a continuous variable.

For example, height, age and weight of people are continuous variables.

b. **Discontinuous (discrete) variable:** A variable which cannot take all possible two given values is called a discontinuous or discrete variable.

For example, the number of members in a family and the number of workers in a factory are discrete variables (since the variable cannot take any value between 1 and 2, 2 and 3 etc.).

5. **Range:**

The difference between the maximum and minimum values of a variable is called its range.

6. **Variate:**

A particular value of a variable is called variate (observation).

7. **Frequently:**

The number of times a variate (observation) occurs in a given data is called frequency of that variate.

8. **Frequency distribution:**

A tabular arrangement of given numerical data showing the frequency of different variates is called Frequency Distribution, and the table itself is called Frequency Distribution table.

9. **Discrete Distribution:**

In this distribution, the upper limit of one class does not coincide with the lower limit of the next class.

10. **Continuous Distribution:**

In this distribution, the upper limit of one class coincide with the lower limit of next class.

**Converting Discrete Distribution to Continuous Distribution:**

To convert discrete classes into continuous classes, we require Adjustment Factor.

$$\text{Adjustment Factor} = \frac{\text{Lower Limit of one class} - \text{Upper limit of previous class}}{2}$$

Subtract the adjustment factor from all the lower limits and add the adjustment factor to all the upper limits.

11. **True Class Limit:**

In a continuous distribution, the class limits are called True Class Limit.

12. **Actual Class Limit:**

In the discrete distribution, the class limits obtained after adjustment are the True or Actual Class Limits. This Actual Class Limits are also called Class Boundaries.

**Note:**

- If the classes are of equal size then class size = difference between two successive class marks.
- In the classes are equal size and 'h' is the size of each class and 'm' is the mid-value of a class, then,
- 

$$\begin{aligned}\text{Lower Limit of the class} &= \text{mid value} - \text{half the width of class} \\ &= m - h/2\end{aligned}$$

$$\begin{aligned}\text{Upper Limit of the class} &= \text{mid value} + \text{half the width of class} \\ &= m + h/2\end{aligned}$$

13. **Cumulative Frequency :**

The sum of frequencies of all the previous classes and that particular class is called Cumulative Frequency of the class.

14. **Class Size:**

The difference between the actual upper limit and the actual lower limit of a class is called its class size.

15. **Class Mark:**

The class Mark of a class is the value midway between its own actual lower limit and actual upper limit.

#### 16. Mean:

Mean of a number of observations is the sum of the values of all the observations divided by the total number of observations.

Mean = Sum of all observations / Total number of observations

#### Methods to find Mean

- a. **Direct Method:** In this case, we assume that the frequency in each class is centred at its class Mark.

If there are  $n$  classes and  $f_i, y_i$  denote the frequency and class marks respectively of  $i^{\text{th}}$  class.

$$\text{Mean} = \frac{\sum f_i y_i}{\sum f_i}$$

- b. **Short Cut Method:** In this method,  
An approximate mean is taken say  $a$   
and Deviation  $d_i = y_i - a$

$$\text{Mean} = a + \frac{\sum f_i d_i}{\sum f_i}$$

- c. **Step deviation Method:** If  $c$  is the width of each class and

$$u_i = \frac{y_i - a}{c}$$

$$\text{Mean} = a + c \times \frac{\sum f_i u_i}{\sum f_i}$$

#### 17. Median:

Median is the central value of a statistical data if it is arranged in ascending or descending order.

$$\text{Median} = \frac{n+1}{2} \text{ (n is odd)}$$

$$\text{Median} = \frac{n/2 \text{ th observation} + (n/2+1) \text{ th observation}}{2} \text{ (n is even)}$$

#### 18. Quartiles:

Quartiles are the values of a statistical data which divide the whole set of observations into four equal parts.

#### 19. Lower Quartile:

If the variates are arranged in ascending order, then the observation lying mid-way between the lower extreme and the median is called the lower quartile.

It is denoted by  $Q_1$ .

$$\text{Lower quartile } (Q_1) = \frac{n+1}{4} \text{ th observation, (if n is odd)}$$

$$\text{Lower quartile } (Q_1) = \frac{n}{4} \text{ th observation, (if n is even)}$$

**20. Upper Quartile:**

If the variates are arranged in ascending order, then the observation lying mid-way between the median and upper extreme is called the upper quartile.

It is denoted by  $Q_3$ .

Upper quartile ( $Q_3$ ) =  $\frac{3(n+1)}{4}$  th observation, (if n is odd)

Upper quartile ( $Q_3$ ) =  $\frac{3n}{4}$  th observation, (if n is even)

**21. Inter quartile Range:**

The difference between the upper quartile ( $Q_3$ ) and the lower quartile ( $Q_1$ ) is called the inter quartile range.

Thus, interquartile range =  $Q_3 - Q_1$

and

Semi interquartile Range =  $\frac{Q_3 - Q_1}{2}$

**22. Mode:**

Mode of a statistical data is the variate which occurs most frequently. In other words, mode of a statistical data is the variate which has maximum frequency.

**23. Model Class:**

The classes with maximum frequency is called the modal class.

- **Relation between Mode, Median and Mean**

**Mode = 3 Median - 2 Mean**

This is known as Empirical Relation.