Education that enlightens!

# Statistics

1. Facts or figures collected with a definite purpose are called **data**.
2. **Statistics** deals with collection, presentation, analysis and interpretation of numerical data.
3. Arranging data in an order to study their salient features is called **presentation of data**.
4. Data arranged in ascending or descending order is called arrayed data or an **array**.
5. When an investigator with a definite plan or design in mind collects data first handedly, it is called **primary data**.
6. Data when collected by someone else, say an agency or an investigator, comes to you, is known as the **secondary data**.
7. **Range** of the data is the difference between the maximum and the minimum values of the observations.
8. The small groups obtained on dividing all the observations are called classes or **class intervals** and the size is called the **class size** or class width.

   Class size = Upper limit – Lower limit
9. **Class mark** of a class is the mid-value of the two limits of that class.
10. The number of times an observation occurs in the data is called the **frequency** of the observation.
11. A **frequency distribution** in which the upper limit of one class differs from the lower limit of the succeeding class is called an Inclusive or discontinuous frequency distribution.
12. A frequency distribution in which the upper limit of one class coincides with the lower limit of the succeeding class is called an exclusive or **continuous frequency distribution**.
13. In case of continuous frequency distribution, the upper limit of a class is not to be included in that class while in discontinuous both the limits are included.
14. The **cumulative frequency** of a class-interval is the sum of frequencies of that class and the classes which precede (come before) it.
15. A data can be represented **graphically** through:

    (i) Bar graph (ii) Histogram (iii) Frequency polygon
16. A **bar graph** is a diagram showing a system of connections or interrelations between two or more things by using bars.
17. In a bar graph, rectangular bars of uniform width are drawn with equal spacing between them on one axis, usually the *x*-axis. The value of the variable is shown on the other axis that is the *y*-axis.

18. A **histogram** is a graphical representation of a frequency distribution in the form of rectangles with class intervals as bases and heights proportional to the corresponding frequencies such that there is no gap between any two successive rectangles.

19. If classes are not of equal width, then the height of the rectangle is calculated by the ratio of the frequency of that class, to the width of that class

20. **Frequency polygons** are a graphical device for understanding the shapes of distributions.

21. If both a histogram and a frequency polygon are to be drawn on the same graph, then we should first draw the histogram and then join the mid-points of the tops of the adjacent rectangles in the histogram with line-segments to get the frequency polygon.

22. A measure of central tendency tries to estimate the central value which represents the entire data.

23. The three **measures of central tendency** for ungrouped data are mean, mode and median.

24. The **mean** value of a variable is defined as the sum of all the values of the variable divided by the number of values.

25. If $x_1, x_2, x_3 \ldots\ldots, x_n$ are $n$ values of a variable X, then the arithmetic mean of these values is given by:

$$\text{Mean } (\bar{x}) = \frac{1}{n}\sum_{i=1}^{n} x_i$$

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

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

26. **Median** is the value of middle most observation(s).

27. The median is calculated only after arranging the data in ascending order or descending order.

$$\text{Median} = \begin{cases} \text{If } n \text{ is odd, then median} = \left(\frac{n+1}{2}\right)^{\text{th}} \text{ observation} \\ \text{If } n \text{ is even, then median} = \dfrac{\left(\frac{n}{2}\right)^{\text{th}} \text{ observation} + \left(\frac{n}{2}+1\right)^{\text{th}} \text{ observation}}{2} \end{cases}$$

28. **Mode** of a statistical data is the value of that variate which has the maximum frequency.

29. The variate corresponding to the highest frequency is to be taken as the mode and not the frequency

30. The disadvantage of arithmetic mean is that it is affected by extreme values.

31. The disadvantage of mode is that it is not uniquely defined in many cases.