1. What is information?
Knowing about something is called information.

2. What is data?
The collection of meaningful information as facts and numerical figures is called data.

Note: The term data handling was first used by Sir Ronald Aylmer Fisher (1890–1962), a pioneer in the field of statistics.

3. Define information handling.
Information handling is the process of collecting, organizing, summarizing, analyzing, and interpreting numerical data.

4. What are the two main types of data?
The two main types of data are:

1. **Discrete Data**: It can take only specific values. Whole numbers are used to represent discrete data. It is obtained only by **counting**.

   For example, number of books sold by a shopkeeper, number of patients visiting a hospital in a week.

2. **Continuous Data**: It can take every possible value in a given interval. Decimal numbers are used to represent continuous data. It is obtained only by **measuring**.

   For example, mass of students in a class (e.g., 28.5 kg, 26.5 kg, 27.5 kg).

5. What is difference between ungrouped data and grouped data?

Ungrouped Data	Grouped Data
Data that is not arranged in any systematic order (groups or classes) is called ungrouped data.	When data is arranged systematically into classes, it is called grouped data.
It is also known as raw data.	Grouped data organizes raw data into intervals for clearer analysis.
Example: 10, 5, 8, 12, 15, 20, 25, 30, ...	Example: Classes: 5–9, 10–14, 15–19, ... with tally marks and frequencies.

6. Define class limits.
The minimum and the maximum values defined for a class or group are called class limits. The minimum value is called the lower-class limit and the maximum value is called the upper-class limit.

7. What is a frequency distribution?
A frequency distribution is a distribution or table that represents classes or groups along with their respective class frequencies.

8. What are the steps to construct a frequency distribution?
Step 1: Calculate the Range: Range is the difference between the greatest value and the smallest value.

$$Range = X_{max} - X_{min}$$

Example: If the greatest value is 136 and the smallest is 30:

$$Range = 136 - 30 = 106$$

Step 2: Determine Class Size: Divide the range by the number of classes or groups you wish to make.

$$Class\ Size = \frac{Range}{Number\ of\ Classes}$$
$$h = \frac{106}{10}$$
$$h = 10.6$$
$$h \approx 11$$

Step 3: Prepare Four Columns:

(i) Class Limits
(ii) Tally Marks
(iii) Frequencies
(iv) Class Boundaries

Step 4: Make Classes Using the Calculated Class Size: Start from the smallest value. Example: 30–40, 41–51, 52–62, and so on.

Step 5: Tally the Data: Place a tally mark (|) for each data point in its class. Use a diagonal strike on every fifth tally for grouping: ||||

9. How are class boundaries usually found?

Class boundaries usually are found by the following method:

1. Choose the upper class limit of the 1st class and the lower class limit of the 2nd class.
2. Find the difference between these two limits.
3. Divide the difference by 2.
4. Subtract this value from the lower class limit and add it to the upper class limit.

Note: Class boundaries may also be obtained from the **midpoints (x)** using the formula:

$$Class\ Boundaries = x \pm \frac{h}{2}$$

where h is the difference between any two consecutive values of x.

10. The following data has been obtain after weighing 40 students of class V. Make the frequency distribution taking class interval size 5. Also find class boundaries, midpoints and less than cumulative frequency distribution. 34, 26, 33, 32, 24, 21, 37, 40, 41, 28, 28, 31, 33, 34, 37, 23, 27, 31, 31, 36, 29, 35, 36, 37, 38, 22, 27, 28, 29, 31, 35, 35, 40, 21, 32, 33, 27, 29, 30, 23

Solution: Let

$$Weights\ of\ students = X$$
$$Minimum\ value = X_{min} = 21$$
$$Maximum\ value = X_{max} = 41$$
$$Size\ of\ class\ interval = h = 5$$

Classes	Tally Marks	Frequency (f)	Midpoints (X)	$C.B$
$20 - 24$	||||		6	22	$19.5 - 24.5$
$25 - 29$	|||| ||||	10	27	$24.5 - 29.5$
$30 - 34$	|||| |||| ||	12	32	$29.5 - 34.5$
$35 - 39$	|||| ||||	9	37	$34.5 - 39.5$
$40 - 44$	|||	3	42	$39.5 - 44.5$

11. Define histogram.

A histogram is a graph of adjacent rectangles constructed on the $xy\text{-}plane$. It is a graph of frequency distribution.

12. Define frequency polygon.

A frequency polygon is a closed geometrical figure displaying a frequency distribution.

13. Define midpoint.

A midpoint is the average value of the lower and upper class limits. Midpoint is also known as the **class mark**. It is calculated by the formula:

$$Midpoint = \frac{Lower\ class\ limit + Upper\ class}{2}$$

14. What is meant by measure of central tendency?

The measure that gives the centre of the data is called measure of central tendency.

Therefore, measure of central tendency is used to find out the middle or central value of a data set. The measures of central tendency are given by

(i) Arithmetic mean **(ii)** Median **(iii)** Mode **(iv)** Weighted Mean

15. Define arithmetic mean.

Arithmetic Mean (A.M.) is defined as the value of a variable which is obtained by dividing the sum of all the values (observations) by the number of observations.

The arithmetic mean of a set of values $x_1, x_2, x_3, \ldots, x_n$ is denoted by $\bar{X}$ (read as "X-bar") and is calculated as:

$$\bar{X} = \frac{x_1 + x_2 + x_3 + \ \ldots + x_n}{n}$$

$$\bar{X} = \frac{\sum X}{n}$$

Arithmetic Mean	
Ungrouped Data	**Grouped Data**
Direct Method $\bar{X} = \frac{Sum\ of\ all\ values\ of\ observation}{no.of\ observations}$ $\bar{X} = \frac{\sum X}{n}$	**Direct Method** $\bar{X} = \frac{\sum fX}{\sum f}$
Indirect Method **(i) Shortcut** $\bar{X} = A + \frac{\sum D}{n}$ $D = X - A$, where A is any assumed value of X called assumed or provisional **(ii) Coding Method** $\bar{X} = A + \frac{\sum u}{n} \times h$ $u = \frac{X-A}{h}$, where A is any assumed value of X called assumed or provisional and h is the class interval size for unequal intervals.	**Indirect Method** **(i) Shortcut** $\bar{X} = A + \frac{\sum fD}{\sum f}$ $D = X - A$, where A is any assumed value of X called assumed or provisional and X denotes the midpoint of class or group. **(ii) Coding Method** $\bar{X} = A + \frac{\sum fu}{\sum f} \times h$ $u = \frac{X-A}{h}$, where A is any assumed value of X called assumed or provisional and h is the size of class interval.

16. What is Median?

Median is the middle most value in an arranged (ascending or descending order) data set. It is the value which divides the data into two equal parts.

Median is denoted by $\tilde{X}$ (read as X-$tilde$).

Median	
Ungrouped Data	**Grouped Data**
Case 1: When the number of observations is odd $\tilde{X} = \left(\frac{n+1}{2}\right)^{th} \ observation$ **Case 2: When the number of observations is even** $\tilde{X} = \frac{1}{2}\left[\left(\frac{n}{2}\right)^{th} observation + \left(\frac{n+2}{2}\right)^{th} observation\right]$	$Median = l + \frac{h}{f}\left[\frac{n}{2} - c\right]$ Where l = *lower class boundary of the median class* h = *The size of class limits of median class* f = *frequency of median class* n = *Total frequency i.e.,* $\sum f$ c = *cumulative frequency of the class preceding the median class* $Median\ class = the\ class\ containing \left(\frac{n}{2}\right)^{th} observation$

17. What is Mode?

In a data set, the value (observation) which appears or occurs most often is called the mode of the data. It is the most common value.

Mode is denoted by $\hat{X}$ (read as X-hat).

Mode	
Ungrouped Data	**Grouped Data**
$Mode = the\ most\ frequent\ observation$	$Mode = l + \frac{(f_m - f_1)}{(f_m - f_1)(f_m - f_2)} \times h$ Where l = *lower class boundary of modal class* h = *class interval size of modal class* f_m = *frequency of modal class* f_1 = *frequency of the class preceding the modal class* f_2 = *frequency of the class following the modal class*

Note $(Mode)$:

- A data set **can have more than one mode** if multiple values occur most often.
- Sometimes, a data set **may not have any mode** if no value repeats.
- **Mode is hard to find** from a frequency distribution table because we don't know the exact values—only how often a class appears.
- So, we **assume the class with the highest frequency** is the **modal class**.

18. What is Weighted Mean?

Arithmetic Mean is used when all the observations are given equal importance or weight, but there are certain situations in which the different observations get different weights.

In this situation, the **weighted mean**, denoted by $\bar{X}_w$ is preferred.

$$\bar{X}_w = \frac{\sum WX}{\sum W}$$