Q. 1 Explain the terms :– (i) Raw data (ii) Attribute and (iii) Variable.

Ans. **(i) Raw data :** The observations on the units in a population or a sample are called statistical data. The statistical data, before any further processing, are called raw data. These observations are concerning one or more characters possessed by the units. These characters are mainly of two types – some of them can only be described and cannot be measured. These are the qualities of the units. Hence they are called qualitative characters. On the other hand, there are some characters which can be measured in some units. They are called quantitative characters.

(ii) Variable and attribute : Although statistics always deals with numerical data, such data may arise in one of the two ways. In some cases the data are numerical to start with e.g. when we record the height for each of a group of men or the number of rooms in each house of a town. In other cases numbers arise only secondarily, when we record the sex of each newborn baby during a month or the language of each book in a library, the data are not numbers initially. We get numbers if, subsequently, we note the number of male babies and that of female babies, or the number of books written in English, the number written in Hindi, the number written in Bengali and so on.

The first type of data arise if we are observing, for each individual of a group, a character which can be expressed in numbers. Such a character will be referred to as a quantitative character or a **variable**. For the second type of data, the character observed (viz the **sex** of a baby or the language in which a book is written) is not expressible in numerical terms. Such a character is, therefore, called a qualitative character or an **attribute**.

Q. 2. Explain the terms - Discrete and Continuous variables.

Ans. There are two principal types of quantitative characters of statistical data. In the first place, the character may take only some isolated values, like the number of letters in a word (word-length) number of petals in a flower, number of members in a family

(family-size) and so on. Such type of variables are called **discontinuous or discrete variables.** Alternatively, the character may conceivably take any value within its range of variation. The height, weight or age of a man, the diameter of a bobbin, the temperature, rainfall or humidity in a region, etc. are variables of this type. In this case, the actual measurements will present a discreteness e.g. when heights are given correct to the nearest cm. or when weights are given correct to the nearest kg. But this discreteness is completely artificial, being due to the limitations of the measuring instrument. Such type of variables are called continuous variables.

Q. 3. Define the terms - Attribute, Variable, Discrete variable, Continuous variable.

Ans. (i) Attribute : A qualitative characteristic like sex, nationality, religion, grade in examination, blood group, beauty, defectiveness of an article produced by a machine, etc. is called as attribute.

(ii) Variable : A quantitative characteristic (which changes) its value like weight of person, examination marks, population of a country, profit of a salesman, etc. is called as variable.

(iii) Discrete variable : A variable taking only particular values is called as discrete variable. e.g No. of students in a class, no. of articles produced by a machine, population of a county, no. of workers in a factory, etc.

(iv) Continuous variable : A variable taking all possible values in a certain range is called as continuous variable. e.g. weight of a person, length of a screw produced by a machine, temperature at a certain place, agricultural production, electricity consumption of a family, speed of a vehicle, etc.

Q. 4 What do you mean by classification ? Discuss the importance of classification in statistical data.

Ans. The entire process of making homogeneous and non-overlapping groups of observations according to similarities is called as classification. The groups so formed are called as class intervals or classes.

For the sake of statistical analysis, the data items are arranged in increasing (or decreasing) order.

However, if there is a huge amount of observations, then merely ordered arrangement is not enough. It does not furnish much useful

information nor does it reduce the bulk of data. Data in this form is difficult to comprehend, analyse and interpret. e.g. analysis of income of 5000 individuals. So it becomes quite essential to condense the data in a suitable form. Classification can be used as a tool to condense the data.

Importance of classification : Classification condenses the data. It omits unnecessary details in the data. It facilitates the comparison with other data. e.g. In case of classification of income of 5000 individuals, we can find the number of individuals below poverty line or income distribution of two countries can be compared. Classification reveals prominent features of the data. e.g. we can find the income group in which majority of families lie. Classification enables further analysis like computation of averages, dispersion, etc.

Q. 5 Explain the term " frequency distribution. " Describe construction of frequency distribution of discrete variable and random variable.

Ans. Quantitative classification refers to the classification of data according to some characteristics that can be measured, such as height, weight, income, sales, profits, production, etc. Such a distribution is known as frequency distribution. In this type of classification, there are two elements – variable and frequency.

The term variable refers to the characteristic that varies in amount or magnitude in a frequency distribution. A variable may be either continuous or discrete. A continuous variable is capable of manifesting every conceivable fractional value within the range of possibilities such as the height or weight of persons or the weight of a product. Thus in a continuous variable, data are obtained by numerical measurements rather than counting. On the other hand, a discrete variable is that which can vary only by finite " jumps " and cannot manifest every conceiveble fractional value. e.g. no. of rooms in a house can only take certain values such as 1, 2, 3, 4, etc. Thus continuous data are obtained through measurements, while discrete data are derived by counting.

(i) Formation of frequency distribution of discrete variable : Procedure : (a) Find the smallest and the largest observation. (b) Prepare first column of all possible values of variables from the smallest to the largest. (c) Consider the observations one by one. Put a tally mark against the value to which it relates in second column. (d) Count the number of tally makrs and place them in the third

column in front of corresponding value. Tally marks facilitate counting.

For example : In a survey of 35 families in a village, the number of children per family was recorded and the following data obtained :

			3	4	5	6
1	0	2		0	2	5
7	2	3	4	6	3	2
8	4	5	12	3	7	8
7	6	5	3	5	4	3
9	7	9	4			

Represent the data in the form of a discrete frequency distribution.

No. of children	Tally marks	Frequency
0	||	2
1	|	1
2	||||	4
3	ⅢⅢ |	6
4	ⅢⅢ	5
5	ⅢⅢ	5
6	|||	3
7	||||	4
8	||	2
9	||	2
10	–	0
11	–	0.
12	|	1
		Total = 35

(ii) Formation of frequency distribution of continuous variable : Procedure : (a) Find the smallest and the largest observation. Calculate the difference between them. This difference is called as the range. (b) Decide the classes, by dividing the range into several intervals. The number of classes be preferably between 7 and 20. (c) Prepare first column of table by entering the class intervals. (d) Classify the observations one by one in the appropriate

class by putting tally marks in the second column against the corresponding class. Cross the observation from the original data to avoid double counting. (e) Count the tally marks and enter the number in the third column.

Further there are two methods of classification of continuous variable : -

Inclusive method : In this method, the observation equal to upper limit is included in the same class. Therefore, the method is called as inclusive method. It can be observed that the upper limit of a class is not the same as the lower limit of succeeding class. Therefore, a discontinuity is observed between the classes.

For example : Classify the following data by taking class-intervals such that their mid-values are 17, 22, 27, 32 and so on.

30, 42, 30, 54, 40, 48, 15, 17, 51, 42, 25, 41, 30, 27, 42, 36, 28, 26, 37, 54, 44, 31, 36, 40, 36, 22, 30, 31, 19, 48, 16, 42, 32, 21, 22, 46, 33, 41, 21.

Since we have to classify the data in such a manner that the mid-values are 17, 22, 27, etc., the first class should be 15 – 19, second class 20 – 24, etc.

Variable	Tally marks	Frequency
15 – 19	IIII	4
20 – 24	IIII	4
25 – 29	IIII	4
30 – 34	NN III	8
35 – 39	IIII	4
40 – 44	NN IIII	9
45 – 49	III	3
50 – 54	III	3
		Total = 39

Exclusive method : In this method, the observation equal to upper limit does not belong to the same class. It is included in the next class. Therefore, the method is called as exclusive method. e.g. the observation 4000 is included in 4000 – 5000. In other words, the observation equal to upper limit is excluded from the same class. In this case upper limit of one class is the lower limit of subsequent class. The classes are observed to be continuous without any gap in between them.

For example : The marks obtained by 50 students are given below :

31	13	46	31	30	45	38	42	30	9
30	30	46	36	2	41	44	18	29	63
44	30	19	5	44	15	7	25	12	30
6	22	24	37	15	6	39	32	21	20
42	31	19	14	23	28	17	53	22	21

Construct a grouped frequency distribution.

The lowest value is 2 and largest is 63. The appropriate class intervals shall be 10 because 7 classes would be formed by taking 10 as class interval.

Marks	Tally marks	No. of students
0 – 10	ⅢⅠ Ⅰ	6
10 – 20	ⅢⅠ ⅠⅠⅠⅠ	9
20 – 30	ⅢⅠ ⅢⅠ	10
30 – 40	ⅢⅠ ⅢⅠ ⅠⅠⅠⅠ	14
40 – 50	ⅢⅠ ⅠⅠⅠⅠ	9
50 – 60	Ⅰ	1
60 – 70	Ⅰ	1
		Total = 50

Q. 6 Explain the terms – (i) Class - limits (ii) class-boundaries (iii) class-mark (iv) class-width (v) open frequency