STATISTICS

INTRODUCTION OF STATISTICSC

INTRODUCTION :

The word ‘statistics’ is derived from the latin word ‘status’ meaning ‘a (political) state’. In its origin, statistics was simply the collection of data on different aspects of the life of people, useful to the state.

Statistics deals with collection, organisation, analysis and interpretation of data. The word ‘statistics’ has different meanings in different contexts.

In the second sentence, the word ‘statistics’ is used as a singular noun, meaning the subject analysis of data as well as drawing of meaningful conclusions from the data.

In this chapter we shall extend the study of these three measures, i.e. mean, median and mode from ungrouped data to that of grouped data and also we shall discuss the concept of cumulative frequency and cumulative frequency distribution ; mean, mode and median of discrete and continuous frequency distribution to draw cumulative frequency curve, called ogive and to find the median by using the ogive

Statistics is basically the study of numerical data. The word statistics is used in two Different senses,

(i) In plural sense statistics means data.

(ii) In singular sense, statistics is the science which deals with the collection, analysis and interpretation numberical data .

SOME BASIC DEFINITIONS :

Data :

The word **data** means information in the form of numerical figures or a set of given facts. It is the collection of facts & figures. For example, the percentage of marks scored by 10 students of a class in a test are : 36, 80, 65, 75, 94, 48, 12, 64, 88 and 98.

The set of these figures is the data related to the marks obtained by 10 students in a class test.

Primary Data :

Fresh or first hand data is called primary data. Investigator collect data himself eg. notes, lists, census study data etc.

Secondary Data :

When investigator does not collect the data himself rather collects the data from other published or unpublished sources. eg. Published reports, official statistics collected by the Government on various facts.

Raw Data :

Data obtained from direct observation is called raw data.

The marks obtained by 10 students in a monthly test is an example of raw data or ungrouped date. So, to make this data clearer and more meaningful, we group it into ordered intervals.

Grouped Data :

To present the data in a more meaningful way, we condense the data into convenient number of classes or groups, generally not exceeding 10 and not less than 5.

Observation :

Each numerical figure in a data is called an observation.

Frequency :

The number of times a particular observation occurs is called its frequency.

Discrete Frequency Distribution :

If each data is given with their frequency, then this type of frequency distribution is called discrete frequency distribution.

Continuous Frequency Distribution :

If the data is given in the form of class interval with frequency, then this type of frequency distribution is called continuous frequency distribution for example

Marks	35-45	45-55	55-65	65-75	75-85	85-95
Frequency	3	4	8	4	5	1

Frequency :

The number of observations in each class is called frequency of that class. In Table the frequency of class 55-65 is 8 and that of the class 85-95 is 1.

Class-Intervals and Class Limits :

In the frequency Table 35-45 is called "class-interval" and the end numbers, 35 and 45 are called "class limits", the smaller number 35 is the lower class limit and the larger number 45 is the upper class limit.

Range :

The difference between the maximum and the minimum value of the given observations is called the range of the data.

Given x_1, x_2 ..... x_n (n individual observations)

Range = (Maximum Value) – (Minimum Value)

Note : The range of the class interval is the difference of the actual limits of the class.

Class Boundaries :

In an exclusive form, the lower and upper limits are known as class boundaries or true lower limit and true upper limit of the class respectively.

Thus, the boundaries of 35–45 in exclusive form 35 and 45.

The boundaries in an inclusive form are obtained by substracting 0.5 to the lower limit and adding 0.5 to the upper limit.

Thus, the boundaries of 35–45 in the inclusive form are 34.5 – 45.5.

The size or width of a Class Interval :

The size or width of a class-interval is the difference between the lower and upper class boundaries.

e.g., size = 44.5 – 34.5 = 10

Class Mark :

Now, for each class-interval, we require a point which would serve as the representative of the whole class. It is assumed that the frequency of each class interval is centred around its mid-point. So the mid-point (or class mark) of each class can be chosen to represent the observations falling in the class. we find the mid-point of a class (or its class mark) by finding the average of its upper and lower limits. That is,

$$\text{Class marks} = \frac{\text{Upper class limit} + \text{Lowet class limit}}{2}$$

For example, class mark of class 35 – 45 $= \frac{35+45}{2} = \frac{80}{2} = 40.$

Cumulative Frequency Table :

The total of frequencies of all the previous and the given class is called the cumulative frequency of the class e.g.

Class-intervals	Frequency	Cumulative frequency
35-45	3	3
45-55	4	7
55-65	8	15
65-75	4	19
75-85	5	24
85-95	1	25
Total	25	

TABULATION OR PRESENTATION OF DATA :

A systematical arrangement of the data in a tabular form is called tabulation or presentation of the data. This grouping results in a table called the frequency table which indicates the number of scores within each group.

The quantitative data that is to be analysed statistically can be divided into three categories:

(1) Individual series

(2) Discrete series and

(3) Continuous series

(1) Individual series :

Any raw data that is not grouped.

Example :

(i) The weights of 5 students : 32, 40, 65, 48 and 54 (in kg)

(ii) Percentage marks obtained by 10 students in a test : 48, 59, 63, 72, 48, 72, 84, 98, 90 and 60

(2) Discrete series :

A discrete series is formulated from raw data by taking the frequency of the observation into consideration.

Example :

Given below is the data showing the number of computers in 12 families of a locality : 1, 1, 2, 3, 2, 1, 4, 3, 2, 2, 1, 1

Arranging the data in the ascending order : 1, 1, 1, 1, 1, 2, 2, 2, 2, 3, 3, 4

We may prepare a frequency table as below :

Number of computers	Number of families (frequency)
1	5
2	4
3	2
4	1

(3) Continuous series :

When the data contains large number of observations, we put them into different groups called class intervals such as 1–10, 11–20, 21–30, etc.

Here, 1 – 10 means data whose values lie between 1 and 10 including both 1 and 10. This form is known as inclusive form. Also, 1 is called the lower limit and 10 is called the upper limit.

Example :

Given below are the marks (out of 50) obtained by 30 students in an examination. Taking class intervals 1–10, 11–20, 21–30, 31–40, and 41–50, we construct a frequency distribution table for the above data.

First, we write the marks in the ascending order as :

Now, we can prepare the frequency distribution table as below.

Class interval	Frequency
1 – 10	4
11 – 20	6
21 – 30	8
31 – 40	5
41 – 50	7

Ex. 1 Find the range of {2, 7, 6, 4, 3, 8, 5, 12}.

Sol. Arranging the given data in the ascending order

We have ; {2, 3, 4, 5, 6, 7, 8, 12}

$\therefore$ Range = (Maximum value) – (Minimum value) = 12 – 2 = 10

Ex. 2 Given below are the ages of 25 students of class IX in a school. Prepare a discrete frequency distribution. 15, 16, 16, 14, 17, 17, 16, 15, 15, 16, 16, 17, 15, 16, 16, 14, 16, 15, 14, 15, 16, 16, 15, 14, 15.

Sol. Frequency distribution of ages of 25 students

Age	Tally marks	Frequency
14	𝍸	4
15	𝍸 | | |	8
16	𝍸 𝍸	10
17	| | |	3
Total		25

Ex.3 Form a discrete frequency distribution from the following scores:-

Sol. 15, 18, 16, 20, 25, 24, 25, 20, 16, 15, 18, 18, 16, 24, 15, 20, 28, 30, 27, 16, 24, 25, 20, 18, 28, 27, 25, 24, 24, 18, 18, 25, 20, 16, 15, 20, 27, 28, 29, 16.

Frequency Distribution of Scores		
Variate	Tally marks	Frequency
15	𝍸	4
16	𝍸 |	6
18	𝍸 |	6
20	𝍸 |	6
24	𝍸	5
25	𝍸	5
27	| | |	3
28	| | |	3
29	|	1
30	|	1
Total		40

Ex.4 The water tax bills (in rupees) of 30 houses in a locality are given below. Construct a grouped frequency distribution with class size of 10.

30, 32, 45, 54, 74, 78, 108, 112, 66, 76, 88, 40, 14, 20, 15, 35, 44, 66, 75, 84, 95, 96, 102, 110, 88, 74, 112, 14, 34, 44.

Sol. Here the maximum and minimum values of the variate are 112 and 14 respectively.

$\therefore$ Range $= 112 - 14 = 98$.

It is given that the class size is 10, and

$$\frac{\text{Range}}{\text{Class size}} = \frac{98}{10} = 9.8$$

So, we shoule have 10 classes each of size 10.

The minimum and maximum values of the variate are 14 and 112 respectively. So we have to make the classes in such a way that first class includes the minimum value and the last class includes the maximum value. If we take the first class as 14-24 it includes the minimum value 14. If the last class is taken as 104-114, then it includes the maximum value 112. Here, we form classes by exclusive method. In the class 14-24, 14 is included but 24 is excluded. Similarly, in other classes, the lower limit is included and the upper limit is excluded.

In the view of above discussion, we construct the frequency distribution table as follows

Bill (in rupees)	Tally marks	Frequency
14-24	||||	4
24-34	||	2
34-44	|||	3
44-54	|||	3
54-64	|	1
64-74	||	2
74-84	𝍸	5
84-94	|||	3
94-104	|||	3
104-114	||||	4
Total		30

Ex.5 The marks obtained by 40 students of class IX in an examination are given below :

18, 8, 12, 6, 8, 16, 12, 5, 23, 2,16, 23, 2, 10, 20, 12, 9, 7, 6, 5, 3, 5, 13, 21, 13, 15, 20, 24, 1, 7, 21, 16, 13, 18, 23, 7, 3, 18, 17, 16.

Present the data in the form of a frequency distribution using the same class size, one such class being 15-20 (where 20 is not included)

Sol. The minimum and maximum marks in the given raw data are 0 and 24 respectively. It is given that 15-20 is one of the class intervals and the class size is same. So, the classes of equal size are 0-5, 5-10, 10-15, 15-20 and 20-25

Thus, the frequency distribution is as given under : Frequency Distribution of Marks

Marks	Tally marks	Frequency
0-5	𝍸𝍷	6
5-10	𝍸𝍷𝍸𝍷	10
10-15	𝍸𝍷 III	8
15-20	𝍸𝍷 III	8
20-25	𝍸𝍷 III	8
	Total	40

CENTRAL TENDENCY :

A no. or quantity which is typical or representative of a set of data is called central tendency. The measure of this kind is known as averages.

Measures of central tendancy or average are usually of the following types