

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. Range of the data is the difference between the maximum and the minimum values of the observations.
6. A variable which can theoretically assume any value between two given values is called a continuous variable. E.g. Height of an individual can be 160 cm or 160.325 cm depending on the accuracy of the measurement.
7. A variable which can only assume countable values or finite values is called a discrete variable. E.g. the number of children in a family can only be 1, 2, 3 etc. and not 1.2, 3.5 or 2.7 etc.
8. Thus, a set of data containing discrete variables is called discrete set of data and the set of data which can be described by a continuous variable is called continuous set of data.
9. The number of times an observation occurs in the data is called the frequency of the observation.
10. A frequency distribution is a tabular arrangement of data showing the frequency with which each successive value of a variable occurs.
11. A grouped frequency distribution is a set of data arranged in groups so that we know the number of times that each of the various values occurs in different groups.
12. Class limits are two figures in which a class is bounded. For e.g. in the class 1 – 5, the two class limits are 1 and 5. 1 is called the lower limit and 5 is called the upper limit.
13. A frequency distribution in which the upper limit of one class differs from the lower limit of the succeeding class is called an Exclusive or discontinuous frequency distribution. For E.g. the below table shows an exclusive frequency distribution.

Marks	No. of Students
1 - 5	7
6 – 10	8
11 – 15	4
21 – 25	2

14. A frequency distribution in which the upper limit of one class coincides from the lower limit of the succeeding class is called an exclusive or continuous frequency distribution. For E.g. the below table shows a exclusive frequency distribution.

Marks	No. of Students
1 – 6	5
6 – 11	1
11 – 16	6
16 – 21	3

15. In case of exclusive frequency distribution, the upper limit of a class is not to be included in that class while in inclusive frequency distribution both the limits are included.

16. The actual or true limits are also called boundaries. The above table can be re-written with true or actual limits as :

Marks	No. of Students
0.5 – 5.5	5
5.5 – 10.5	1
10.5 – 15.5	6
15.5 – 20.5	3

17. The above table with true limits can only be written for continuous distributions. In case of discontinuous distributions, the stated limits are the actual limits.
18. The difference between the actual lower limit of any class and the actual lower limit of the next larger class is called the class interval of the former class. For e.g. in the above table the class interval of class (0.5 – 5.5) is $5.5 - 0.5 = 5$.
19. The class interval can also be found by subtracting the lower limit from the actual upper limit of the same class. However, this is true only for exclusive (continuous) frequency distribution table. For an inclusive distribution this method does not work.

20. To convert inclusive classes of a frequency distribution table into exclusive classes, find the adjustment factor and add it to the upper limits and subtract it from the lower limits. For e.g. for the below table:

Marks	No. Of Students
5 – 9	3
10 - 14	7

- I. To change the inclusive classes to exclusive we first find the adjustment factor as,

$$\frac{1}{2}(\text{lower limit of second class} - \text{upper limit of first class})$$
- II. Using this we get, $\frac{10-9}{2} = 0.5$, which is the adjustment factor. Subtracting 0.5 from the lower limits and adding to upper limits of the classes in the above table we get,

Marks (Inclusive classes)	Marks (Exclusive classes)	No. of students
5 - 9	4.5 – 9.5	3
10 - 14	9.5 – 14.5	7

21. If the distribution is in exclusive form, then the stated limits will be taken as actual limits.
22. The value midway between the actual lower and actual upper limits of the classes or the stated lower and upper limits of the classes is called the Class Mid-Point or mid-mark or Class Mark.
23. To find the class boundaries using the given class marks we proceed as follows:
- I. Find the class size.
 - II. Divide the class size by 2.
 - III. Subtract and add the result to each class mark to get the corresponding class interval.
24. The following formula is used to determine the number of class intervals,

$$\text{No. of class intervals} = \frac{\text{Range}}{\text{Size of the class interval}} + 1$$

Histogram and Frequency Polygon

1. A data can be represented graphically through:
(i) Bar graph (ii) Histogram (iii) Frequency polygon
2. A bar graph is a diagram showing a system of connections or interrelations between two or more things by using bars.
3. 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.
4. Frequency polygons are a graphical device for understanding the shapes of distributions.
5. 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.
6. 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.