What is Frequency Distribution?

Frequency distribution is used to organize the collected data in table form. The data could be marks scored by students, temperatures of different towns, points scored in a volleyball match, etc. After data collection, we have to show data in a meaningful manner for better understanding. Organize the data in such a way that all its features are summarized in a table. This is known as frequency distribution.

Let's consider an example to understand this better. The following are the scores of 10 students in the G.K. quiz released by Mr. Chris 15, 17, 20, 15, 20, 17, 17, 14, 14, 20. Let's represent this data in frequency distribution and find out the number of students who got the same marks.

Quiz Marks	No. of Students
15	2
17	3
20	3
14	2

We can see that all the collected data is organized under the column quiz marks and the number of students. This makes it easier to understand the given information and we can see that the number of students who obtained the same marks. Thus, frequency distribution in statistics helps us to organize the data in an easy way to understand its features at a glance.

Frequency Distribution Graphs

There is another way to show data that is in the form of graphs and it can be done by using a frequency distribution graph. The graphs help us to understand the collected data in an easy way. The graphical representation of a frequency distribution can be shown using the following:

- Bar Graphs: Bar graphs represent data using rectangular bars of uniform width along with equal spacing between the rectangular bars.
- Histograms: A histogram is a graphical presentation of data using rectangular bars of different heights. In a histogram, there is no space between the rectangular bars.
- Pie Chart: A pie chart is a type of graph that visually displays data in a circular chart. It records data in a circular manner and then it is further divided into sectors that show a particular part of data out of the whole part.
- Frequency Polygon: A frequency polygon is drawn by joining the mid-points of the bars in a histogram.

Types of Frequency Distribution

There are four types of frequency distribution under statistics which are explained below:

- **Ungrouped frequency distribution:** It shows the frequency of an item in each separate data value rather than groups of data values.
- **Grouped frequency distribution:** In this type, the data is arranged and separated into groups called class intervals. The frequency of data belonging to each class interval is noted in a frequency distribution table. The grouped frequency table shows the distribution of frequencies in class intervals.
- Relative frequency **distribution:** It tells the proportion of the total number of observations associated with each category.
- Cumulative frequency **distribution:** It is the sum of the first frequency and all frequencies below it in a frequency distribution. You have to add a value with the next value then add the sum with the next value again and so on till the last. The last cumulative frequency will be the total sum of all frequencies.

Frequency Distribution Table

A frequency distribution table is a chart that shows the frequency of each of the items in a data set. Let's consider an example to understand how to make a frequency distribution table using tally marks. A jar containing beads of different colors- red, green, blue, black, red, green, blue, yellow, red, red, green, green, green, yellow, red, green, yellow. To know the exact number of beads of each particular color, we need to classify the beads into categories. An easy way to find the number of beads of each color is to use tally marks. Pick the beads one by one and enter the tally marks in the respective row and column. Then, indicate the frequency for each item in the table.

Categories	Tally Marks	Frequency
Red	ЖЖ	5
Green	ЖЖ I	6
Blue	II	2
Black	I	1
Yellow	III	3

Thus, the table so obtained is called a frequency distribution table.

Types of Frequency Distribution Table

There are two types of frequency distribution tables: Grouped and ungrouped frequency distribution tables.

Grouped Frequency Distribution Table: To arrange a large number of observations or data, we use grouped frequency distribution table. In this, we form class intervals to tally the frequency for the data that belongs to that particular class interval.

For example, Marks obtained by 20 students in the test are as follows. 5, 10, 20, 15, 5, 20, 20, 15, 15, 15, 10, 10, 10, 20, 15, 5, 18, 18, 18, 18. To arrange the data in grouped table we have to make class intervals. Thus, we will make class intervals of marks like 0 – 5, 6 – 10, and so on. Given below table shows two columns one is of class intervals (marks obtained in test) and the second is of frequency (no. of students). In this, we have not used tally marks as we counted the marks directly.

Marks obtained in Test (class intervals)	No. of Students (Frequency)
0 – 5	3
6 – 10	4
11 – 15	5
16 – 20	8
Total	20

Ungrouped Frequency Distribution Table: In the ungrouped frequency distribution table, we don't make class intervals, we write the accurate frequency of individual data. Considering the above example, the ungrouped table will be like this. Given below table shows two columns: one is of marks obtained in the test and the second is of frequency (no. of students).

Marks obtained in Test	No. of Students
5	3
10	4
15	5
18	4
20	4
Total	20

Frequency Distribution Examples

- **Example 1: There are 20 students in a class. The teacher, Ms. Jolly, asked the students to tell their favorite subject. The results are as follows - Mathematics, English, Science, Science, Mathematics, Science, English, Art, Mathematics, Mathematics, Science, Art, Art, Science, Mathematics, Art, Mathematics, English, English, Mathematics.**

  Represent this data in the form of frequency distribution and identify the most-liked subject?

  Solution: 20 students have indicated their choices of preferred subjects. Let us represent this data using tally marks. The tally marks are showing the frequency of each subject.

Subject	Tally Marks	Number of Students
Art	IIII	4
Mathematics	IIII II	7
Science	IIII	5
English	IIII	4

  According to the above frequency distribution, mathematics is the most liked subject.

- **Example 2: 100 schools decided to plant 100 tree saplings in their gardens on world environment day. Represent the given data in the form of frequency distribution and find the number of schools that are able to plant 50% of the plants or more? 95, 67, 28, 32, 65, 65, 69, 33, 98, 96, 76, 42, 32, 38, 42, 40, 40, 69, 95, 92, 75, 83, 76, 83, 85, 62, 37, 65, 63, 42, 89, 65, 73, 81, 49, 52, 64, 76, 83, 92, 93, 68, 52, 79, 81, 83, 59, 82, 75, 82, 86, 90, 44, 62, 31, 36, 38, 42, 39, 83, 87, 56, 58, 23, 35, 76, 83, 85, 30, 68, 69, 83, 86, 43, 45, 39, 83, 75, 66, 83, 92, 75, 89, 66, 91, 27, 88, 89, 93, 42, 53, 69, 90, 55, 66, 49, 52, 83, 34, 36**

  Solution: To include all the observations in groups, we will create various groups of equal intervals. These intervals are called class intervals. In the frequency distribution, the number of plants survived is showing the class intervals, tally marks are showing frequency, and the number of schools is the frequency in numbers.

Number of plants survived	Tally Marks	Number of schools (frequency)
20 – 29	III	3
30 – 39	JHT JHT IIII	14
40 – 49	JHT JHT II	12
50 – 59	JHT III	8
60 – 69	JHT JHT JHT III	18
70 – 79	JHT JHT	10
80 – 89	JHT JHT JHT JHT III	23
90 – 99	JHT JHT II	12
Total		100

So, according to class intervals starting from 50 – 59 to 90 – 99, the frequency of schools able to retain 50% or more plants are 8 + 18 + 10 + 23 + 12 = 71 schools. Thus, 71 schools are able to retain 50% or more plants in their garden.

Cumulative and Relative frequency distribution:

Cumulative Frequency

Cumulative frequency is the total of a frequency and all frequencies in a frequency distribution until a certain defined class interval. The running total of frequencies starting from the first frequency till the end frequency is the cumulative frequency. The total and the data are shown in the form of a table where the frequencies are divided according to class intervals. Let us learn more about cumulative frequency, plotting a cumulative frequency graph, and learn to read a cumulative frequency table along with solving examples.

Definition of Cumulative Frequency

In statistics, the frequency of the first-class interval is added to the frequency of the second class, and this sum is added to the third class and so on then, frequencies that are obtained this way are known as cumulative frequency (c.f.). A table that displays the cumulative frequencies that are distributed over various classes is called a cumulative frequency distribution or cumulative frequency table. There are two types of cumulative frequency - lesser than type and greater than type. Cumulative frequency is used to know the number of observations that lie above (or below) a particular frequency in a given data set. Let us look at a few examples that are used in many real-world situations.

Example 1: Robert is the sales manager of a toy company. On checking his quarterly sales record, he can observe that by the month of April, a total of 83 toy cars were sold.

Month	Number of toy cars sold (Frequency)	Total number of toy cars sold (Cumulative Frequency)
January	20	20
February	30	20 + 30 = 50
March	15	50 + 15 = 65
April	18	65 + 18 = 83

Note how the last cumulative total will always be equal to the total for all observations since all frequencies will already have been added to the previous total. Here, 83=20+30+15+1883=20+30+15+18

Example 2: A Major League Baseball team records its home runs in the 2020 session as given below.

Match	f (home runs)	cf (cumulative total)
Qualifying match	11	11
Quarterfinal match	8	11 + 8 = 19
Semifinal	10	19 + 10 = 29
Final	7	29 + 7 = 36

From the above table, it can be observed that the team made 29 home runs before playing in the finals.

Types of Cumulative Frequency

Cumulative frequency is the total frequencies showcased in the form of a table distributed in class intervals. There are two types of cumulative frequency i.e. lesser than and greater than, let us learn more about both types.

Lesser Than Cumulative Frequency

Lesser than cumulative frequency is obtained by adding successively the frequencies of all the previous classes including the class against which it is written. The cumulate starts from the lowest to the highest size. In other words, when the number of observations is less than the upper boundary of a class that's when it is called lesser than cumulative frequency.

Greater Than Cumulative Frequency

Greater than cumulative frequency is obtained by finding the cumulative total of frequencies starting from the highest to the lowest class. It is also called more than type cumulative frequency. In other words, when the number of observations is more than or equal to the lower boundary of the class that's when it is called greater than cumulative frequency.

Let us look at example to understand the two types.

Example: Write down less than type cumulative frequency and greater than type cumulative frequency for the following data.

Height (in cm)		Frequency (students)
140	–	145	2
145	–	150	5
150	–	155	3
155	–	160	4
160 – 165	1

Solution: We would have less than type and more than type frequencies as:

Height (in inches)	f	Height less than type (upper limits)	c.f	Height less than type (lower limits)	c.f
140 - 145	2	145	2	140	15
145 - 150	5	150	7	145	13
150 - 155	3	155	10	150	8
150 - 160	4	160	14	155	4
160 - 165	1	165	15	160	1

The following information can be gained from either graph or table
- Out of a total of 15 students, 8 students have a height of more than 150 cm
- None of the students are taller than 165 cm
- Only one of the 15 students has a height of more than 160 cm

Constructing a Cumulative Frequency Distribution Table

A cumulative frequency table is a simple visual representation of the cumulative frequencies for different values or categories. To construct a cumulative frequency distribution table, there are a few steps that can be followed which makes it simple to construct. Let us see what the steps are:

- **Step 1:** Use the continuous variables to set up a frequency distribution table using a suitable class length.
- **Step 2:** Find the frequency for each class interval.
- **Step 3:** Locate the endpoint for each class interval (upper limit or lower limit).
- **Step 4:** Calculate the cumulative frequency by adding the numbers in the frequency column.
- **Step 5:** Record all results in the table.

Example: During a 20-day long skiing competition, the snow depth at Snow Mountain was measured (to the nearest cm) for each of the 20 days. The records are as follows: 301, 312, 319, 354, 359, 345, 348, 341, 347, 344, 349, 350, 325,323, 324, 328,322, 332, 334, 337.

Solution:

Given measurements of snow depths are: 301, 312, 319, 354, 359, 345, 348, 341, 347, 344, 349, 350, 325,323, 324, 328,322, 332, 334, 337

Step 1: The snow depth measurements range from 301 cm to 359 cm. To produce the frequency distribution table, the data can be grouped in class intervals of 10 cm each.

In the Snow depth column, each 10-cm class interval from 300 cm to 360 cm is listed.

Step 2: The frequency column will record the number of observations that fall within a particular interval. The tally column will represent the observations only in numerical form.

Step 3: The endpoint is the highest number in the interval, regardless of the actual value of each observation.

For example, in the class interval of 311-320, the actual value of the two observations is 312 and 319. But, instead of using 219, the endpoint of 320 is used.

Step 4: The cumulative frequency column lists the total of each frequency added to its predecessor.

Using the same steps mentioned above, a cumulative frequency distribution table can be made as:

Snow Depth (x)	Tally	Frequency (f)	Endpoint	c.f
300 - 310	I	1	310	1
311 - 320	II	2	320	3
321 - 330	IIIII	5	330	8
331 - 340	III	3	340	11
341 - 350	IIIII II	7	350	18
351 - 360	II	2	360	20

Constructing Cumulative Frequency Distribution Graph

The cumulative frequency distribution of grouped data can be represented on a graph. Such a representative graph is called a cumulative frequency curve or an ogive. Representing cumulative frequency data on a graph is the most efficient way to understand the data and derive results. In the world of statistics, graphs, in particular, are very important, as they help us to visualize the data and understand it better. So let us learn about the graphical representation of the cumulative frequency. There are two types of Cumulative Frequency

Curves (or Ogives): More than type Cumulative Frequency Curve and Less than type Cumulative Frequency Curve.

More Than Cumulative Frequency Curve
In the more than cumulative frequency curve or ogive, we use the lower limit of the class to plot a curve on the graph. The curve or ogive is constructed by subtracting the total from first-class frequency, then the second class frequency, and so on. The upward cumulation result is more than or greater than the cumulative curve. The steps to plot a more than curve or ogive are:
- **Step 1:** Mark the lower limit on the x-axis
- **Step 2:** Mark the cumulative frequency on the y-axis.
- **Step 3:** Plot the points (x,y) using lower limits (x) and their corresponding Cumulative frequency (y).
- **Step 4:** Join the points by a smooth freehand curve.

Less Than Cumulative Frequency Curve
In the mess than cumulative frequency curve or ogive, we use the upper limit of the class to plot a curve on the graph. The curve or ogive is constructed by adding the first-class frequency to the second class frequency to the third class frequency, and so on. The downward cumulation result is less than the cumulative frequency curve. The steps to plot a less than cumulative frequency curve or ogive are:
- **Step 1:** Mark the upper limit on the x-axis
- **Step 2:** Mark the cumulative frequency on the y-axis.
- **Step 3:** Plot the points (x,y) using upper limits (x) and their corresponding Cumulative frequency (y).
- **Step 4:** Join the points by a smooth freehand curve.

Example: Graph the two ogives for the following frequency distribution of the weekly wages of the given number of workers.

Weekly wages	No. of workers
0-20	4
20-40	5
40-60	6
60-80	3

Solution:

Weekly wages	No. of workers	C.F. (Less than)	C.F. (More than)
0-20	4	4	18 (total)
20-40	5	9 (4 + 5)	14 (18 - 4)
40-60	6	15 (9 + 6)	9 (14 - 5)
60-80	3	18 (15 + 3)	3 (9 - 6)

Less than curve or ogive:
Mark the upper limits of class intervals on the x-axis and take the less than type cumulative frequencies on the y-axis. For plotting less than type curve, points (20,4), (40,9), (60,15), and (80,18) are plotted on the graph and these are joined by freehand to obtain the less than ogive.