

UNIT 2

Data Summarization

Data summarization is the first step in **statistics**, it is aimed at extracting useful information. Summary statistics are used to summarize a set of observations, to communicate the largest amount of information as simply as possible.

Data can be summarized **numerically as a table** (tabular summarization), or **visually as a graph** (data visualization).

Frequency Distribution

Frequency is how often something repeats, and a **frequency distribution** is a representation, either in a graphical or tabular format, that **displays the number of observations** within a given interval. It gives a visual display of the frequency of items or shows the number of times they occurred.

Example 1

Tally marks are often used to make a frequency distribution table. For example, let's say you survey a number of households and find out how many pets they own. The results are 3, 0, 1, 4, 4, 1, 2, 0, 2, 2, 0, 2, 0, 1, 3, 1, 2, 1, 1, 3. Looking at that string of numbers boggles the eye; a frequency distribution table will make the data easier to understand.

Number of Pets (x)	Tally	Frequency (f)
0		4
1	I	6
2		5
3		3
4		2

Types of frequency distribution

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.

Steps for constructing Frequency distribution

- Sort the data in ascending order
- Calculate the range of data
- Decide on the number of intervals in the frequency distribution
- Determine the intervals.
- Decide the starting point
- Tally and count the observations under each interval.

Exercise:

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?

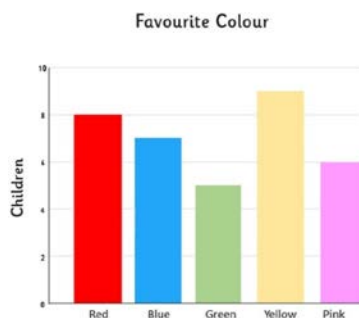
95, 67, 28, 32, 65, 65, 69, 33, 98, 96, 76, 42, 32, 38, 42, 40, 40, 69, 95, 92, 75, 83, 76, 83, 85, 62, 37, 65, 63, 42, 89, 65, 73, 81, 49, 52, 64, 76, 83, 92, 93, 68, 52, 79, 81, 83, 59, 82, 75, 82, 86, 90, 44, 62, 31, 36, 38, 42, 39, 83, 87, 56, 58, 23, 35, 76, 83, 85, 30, 68, 69, 83, 86, 43, 45, 39, 83, 75, 66, 83, 92, 75, 89, 66, 91, 27, 88, 89, 93, 42, 53, 69, 90, 55, 66, 49, 52, 83, 34, 36

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 Graph:

A bar chart or bar graph is a chart or graph that presents categorical data with rectangular bars with heights or lengths proportional to the values that they represent. The bars can be plotted vertically or horizontally. A vertical bar chart is sometimes called a column chart.

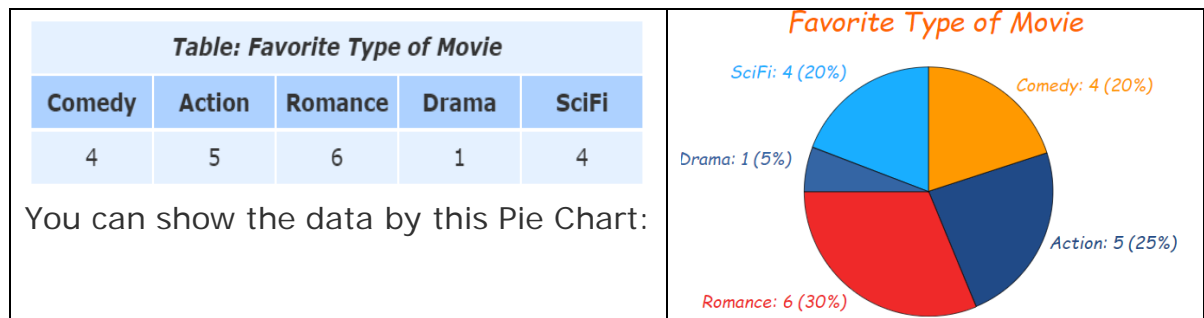


Pie Chart:

A pie chart is a circular statistical graphic, which is divided into slices to illustrate numerical proportion. Or

A Pie Chart is a type of graph that displays data in a circular graph. The pieces of the graph are proportional to the fraction of the whole in each category. In other words, each slice of the pie is relative to the size of that category in the group as a whole. The entire “pie” represents 100 percent of a whole, while the pie “slices” represent portions of the whole.

Imagine you survey your friends to find the kind of movie they like best:



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.

A two-dimensional graphical representation of a continuous frequency distribution is called a histogram. In histogram, the bars are placed continuously side by side with no gap between adjacent bars. That is, in histogram rectangles are erected on the class intervals of the distribution. The areas of rectangle are proportional to the frequencies.

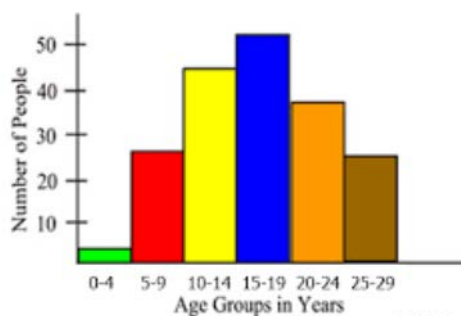
Steps of constructing Histogram:

Step 1 : Represent the data in the continuous form if it is in the discontinuous form.

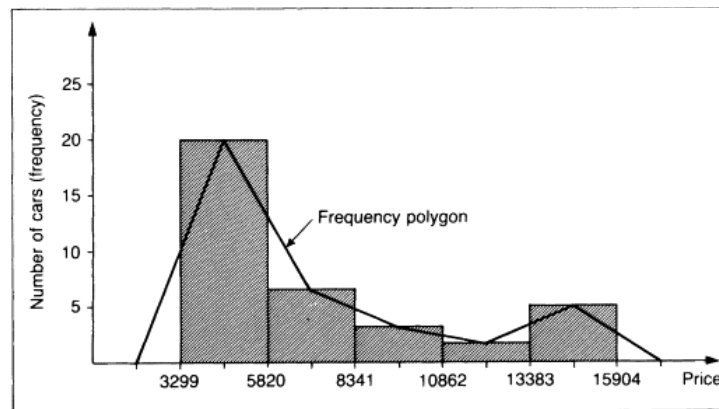
Step 2 : Mark the class intervals along the X-axis on a uniform scale.

Step 3 : Mark the frequencies/Frequency densities along the Y-axis on a uniform scale.

Step 4 : Construct rectangles with class intervals as bases and corresponding frequencies/f.d. as heights.



Frequency Polygon: A frequency polygon is drawn by joining the mid-points of the bars in a histogram.



Cumulative Frequency Polygon (Ogive curve):

A curve that represents the cumulative frequency distribution of grouped data on a 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.

Birth Weight (kg)	2.0-2.5	2.5-3.0	3.0-3.5	3.5-4.0	4.0-4.5	4.5-5.0
Frequency	12	22	33	27	18	8
Cumulative Frequency	12	34	67	94	112	120

