

Data Representation

After the data have been collected, the next step is to present them in some suitable form. The need for proper presentation arises because of the fact that statistical data in their raw form almost defy comprehension. When data are present in easy to- read form, it can help the reader to acquire knowledge in much shorter period of time and also facilitate statistical analysis.

Data representation:

- i) Classification
- ii) Tabulation
- iii) Graphical representation

i) Classification of data:

After collection and editing of data an important step toward processing the data is classification. Classification is the grouping of related facts into different classes. Fact in one class differs from those of another class with respect to some characteristics called a basis of classification.

Objectives of classification:

The process of classification gives prominence to important information gathered while dropping unnecessary details facilitates comparison and enables a statistical treatment of the material collected.

Types of classification:

The data may be classified on the four basis:

- i) Geographical. i.e. area-wise, e.g., cities, district, etc.
- ii) Chronological, i.e., on the basis of time
- iii) Qualitative, i.e. according to some attributes.
- iv) Quantitative, i.e. in terms of magnitude.

ii) Tabulation:

Tabulation involves the orderly and systematic presentation of numeric data in a form designed to elucidate the problem under consideration.

Frequency:

Frequency is in general, the number of occurrences of the items.

Frequency Distribution:

The grouping of data into mutually exclusive classes showing the number of observation in each class is known as frequency distribution.

Class Limit:

Class limits are the highest and lowest values that can be included in the class.

Class interval:

The difference between the upper and lower limit is known as class interval.

Class mid-point = $(\text{upper limit} + \text{lower limit})/2$

Construction of a frequency distribution table:

There are six steps:

Step-1: Select a title for the table.

Step-2: Decide on the number of classes or groups.

We can use 2^k rule.

We will take that value of k as the number of group for which $2^k > n$.

Step-3: Determine the class interval or class width.

$$i \geq \frac{H - L}{k}$$

where,

H = Highest Value

L = Lowest value

k = Number of groups

i = Class interval

Step-4: Set the individual class limit.

Step-5: Use tally marks (|) for each number.

Step-6: Count the number of items in each class.

Problem 1:

The profit in (lakh of taka) of 50 companies for the year 20017-18 are given below:

66, 62, 45, 79, 32, 51, 56, 60, 51, 49, 25, 42, 54, 54, 58, 70, 43, 58, 52, 50, 38, 67,
50, 59, 48, 65, 71, 30, 46, 55, 82, 51, 63, 45, 53, 40, 35, 56, 70, 52, 67, 55, 57, 30, 63, 42, 74, 58, 44, 55

Construct a complete frequency distribution using 5-step procedure/ Arrange the data set through a proper method.

Solution:

Step-1: Title for the table is- “Frequency distribution of profits”.

Step-2: Decide on the number of classes or groups.

When $k=1$, then $2^k=2$

$k=2$, then $2^2=4$

$k=3$, then $2^3=8$

$k=4$, then $2^4=16$

$k=5$, then $2^5=32$

$k=6$, then $2^6=64$

Here, number of observations, $n=50$

So, $k=6$ is the value for which $2^6=64>50$.

Step-3: Determine the class interval or class width.

$$i \geq \frac{H-L}{k} = \frac{82-25}{6} = 9.5 \approx 10$$

H = Highest Value=82

L = Lowest value=25

k = Number of groups=6

i = Class interval

Step-4: Set the individual class limit.

25-35, 35-45, 45-55, 55-65, 65-75, 75-85

Step-5: Use tally marks (|) for each number.

Step-6: Count the number of item in each class.

Frequency distribution of profits

Profits in lakh (class interval)	Tally marks()	Number of companies (Frequency)	Cumulative frequency	Relative Frequency
25-35		4	4	0.08
35-45		7	11	0.14
45-55		15	26	0.3

Most. Salma Yeasmin Jannaty

Lecturer (Statistics), MNS Department,

Brac University

E-mail: salma.yeasmin@bracu.ac.bd

55-65	//// //	14	40	0.28
65-75	///	8	48	0.16
75-85	//	2	50	0.04

N=50

Problem 2: The following data specifies the life of 40 similar car batteries recorded to the nearest tenth of a year:

2.2, 4.1, 3.5, 4.5, 3.2, 3.7, 3.0, 2.6,
3.4, 1.6, 3.1, 3.3, 3.8, 3.1, 4.7, 3.7,
2.5, 4.3, 3.4, 3.6, 2.9, 3.3, 3.9, 3.1,
3.3, 3.1, 3.7, 4.4, 3.2, 4.1, 1.9, 3.4,
4.7, 3.8, 3.2, 2.6, 3.9, 3.0, 4.2, 3.5

Construct a complete frequency distribution / relative frequency distribution table.

iii) **Graphical Representation:**

Data can be presented in various diagrammatic forms. Most popular ones are-

- Bar diagram
- Pie chart
- Histogram
- Frequency polygon
- Percentage polygon
- Cumulative frequency polygon etc.

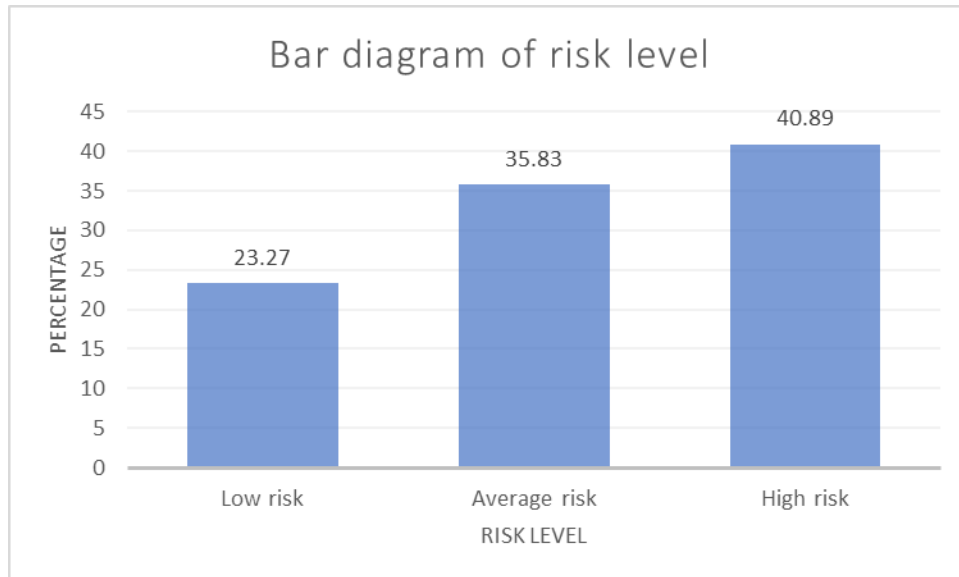
Bar diagram:

- Constructing a bar diagram, the class boundaries of the variables are measured along the x-axis and frequencies are represented along the y-axis.
- Draw bars of equal breadths, the magnitude of each quantity is to be presented by the height of the bar.
- Leave equal space between the bars
- All the bars should have same common base line.

Example:

Frequency and percentage summary table of risk level for 868 mutual funds

Fund risk	Number of funds	Percentage
Low risk	202	23.27
Average risk	311	35.83
High risk	355	40.89



Pie Chart:

To construct a pie-chart a circle is drawn at first which has four right angles at the center and divided into angles (k^0) proportional to the frequency of different categories.

Example:

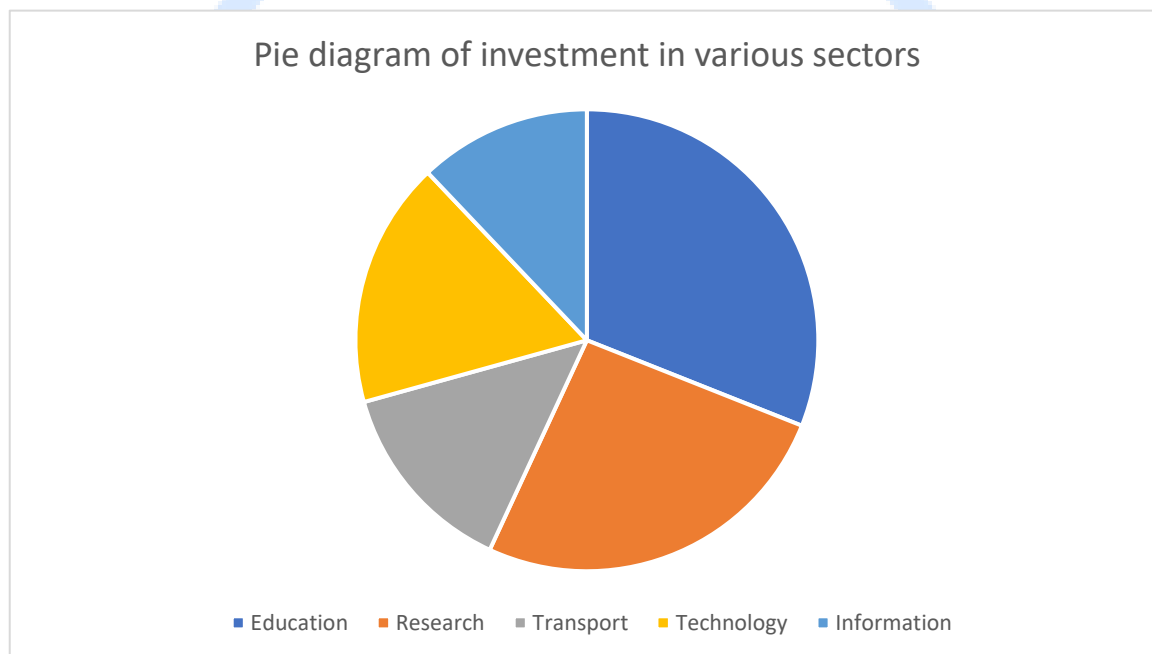
Draw a Pie diagram based on the following information:

Sectors	Amount in Lac Taka
Education	180
Research	150
Transport	80
Technology	100
Information	70

Most. Salma Yeasmin Jannaty
Lecturer (Statistics), MNS Department,
Brac University
E-mail: salma.yeasmin@bracu.ac.bd

Solution:

Sectors	Amount in Lac Taka	Percentage
Education	180	31%
Research	150	26%
Transport	80	14%
Technology	100	17%
Information	70	12%
Total:	580	100



Histogram (continuous):

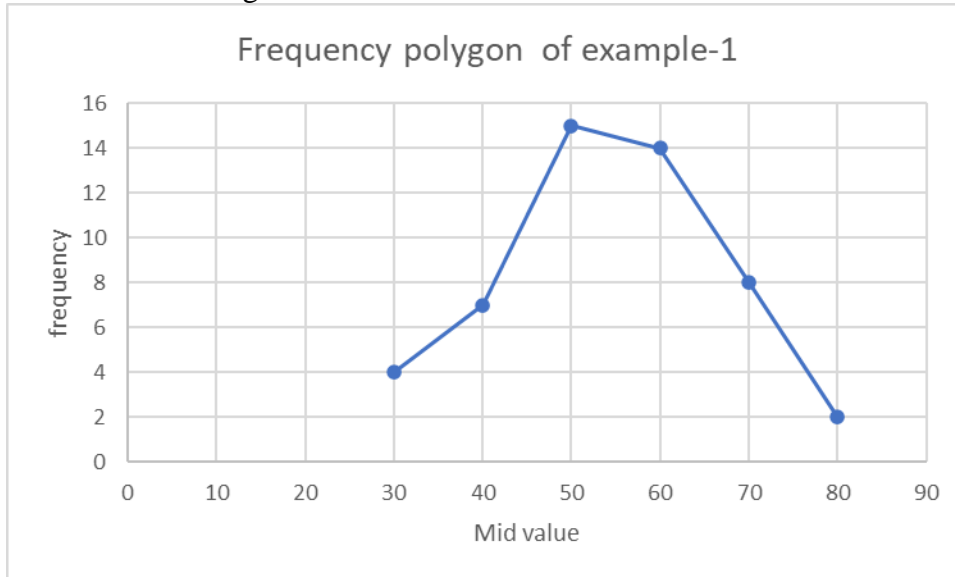
The graph consists a series of rectangles, each of which represents the frequency of scores in one of the class intervals of the frequency distribution table is called histogram.

Construction: Constructing a histogram, the class boundaries of the variables are measured along x- axis and frequencies are presented along y-axis. The vertical bars are drawn on each class boundary according to the corresponding frequencies.

Example of histogram has been given in class lecture.

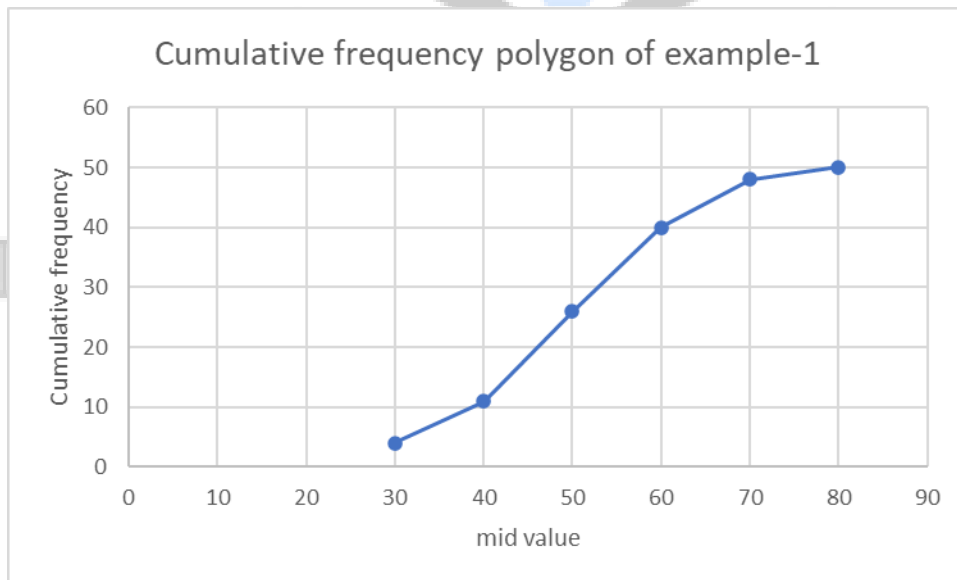
Frequency polygon

Frequency polygon can be constructed by adding the top midpoints on each class intervals of a histogram.



Cumulative Frequency Polygon(Ogive):

We start with the lower limits of the classes and from the total frequencies we subtract the frequency of each class. When these frequencies are plotted, we get a declining curve.





Inspiring Excellence

Most. Salma Yeasmin Jannaty
Lecturer (Statistics), MNS Department,
Brac University
E-mail: salma.yeasmin@bracu.ac.bd