Frequency distribution:

The data that have been collected in original form are called **raw data**. These raw data do not furnish any useful information and thus, they need to be organized in such a way that is easily comprehensible. An appropriate way is to organize them by constructing a **frequency distribution**. The following example illustrates a frequency distribution.

Example 1:

A teacher gave a test in basic science to his Form II students. Their marks were

3	8	6	5	6	4	7	6
5	3	5	6	3	5	4	4
3	6	7	8	1	10	7	6
4	5	0	7	6	5	6	7
1	7	5	4	5	8	5	7

The marks could be tabulated as follows:

Frequency Distribution Table

Mark (x)	**Tally mark**	**Frequency** (f)
0	/	1
1	//	2
2		0
3	////	4
4	~~////~~	5
5	~~////~~////	9
6	~~////~~///	8
7	~~////~~//	7
8	///	3
9		0
10	/	1
		$\sum f = 40$

Now it is easier to gather the following types of information from the frequency table than the raw data.

(a) The highest mark is 10 and the lowest is 0.

(b) 4 students scored more than 7.

(c) 12 students scored less than 5.

(d) 8 students scored 6.

(e) Noone scored 9.

(f) 40 students did the test.

Notes:

(i) The distribution of the variable (i.e. the characteristics under study e.g. marks in the preceding example) along with their frequency is known as **frequency distribution**.

(ii) The **frequency** is the number of times a score (e.g. marks as in the example) is repeated.

(iii) It is sometimes important to compute the **relative frequency distribution**. Realtive frequency of a score is the number of times that a score is made relative to the total number of scores made. It is computed as follows;

$$\text{Relative frquency of a score} = \frac{\text{Frequency of the score}}{\text{Total frequency}} = \frac{f}{\sum f}$$

The relative frequency distribution of Example 1 is given below:

Relative Frequency Distribution

Marks (x)	**Frequency** (f)	**Relative frequency**
0	1	1/40=.025
1	2	2/40=.050
2	0	0/40=.000
3	4	=.100
4	5	=.125
5	9	=.225
6	8	=.200
7	7	=.175
8	3	=.075
9	0	=.000
10	1	=.025
	$\sum f = 40$	

From this table we see that the relative frequency of mark 4 is 0.125. It means that 12.5% (percentage distribution) of the students scored 4 marks.

Grouped Frequency Distribution

If the number of distinct data values is large, the data must be grouped to make them more comprehensible as follows:

We divide an interval containing all the data into a small number of segments, usually of equal width. These segments are called **classes** (or **class intervals**). For example the weights (in kg.) of 50 pieces of luggage are presented in a grouped frequency distribution with the class interval as follows:

Weights (kgs)	**No. of pieces**
7–9	2
10–12	8
13–15	14
16–18	19
19–21	7
	50

From the preceding frequency distribution table we observe the followings:

(i) The intervals of weights i.e. $7-9, 10-12, \ldots, 19-21$ are known as **class intervals**.

(ii) $7, 10, \ldots, 19$ are called **lower limits** of the respective classes.

(iii) $9, 12, \ldots, 21$ are called **upper limits** of the respective classes.

(iv) 6.5–9.5, 9.5–12.5, 12.5–15.5, 15.5–18.5 and 18.5–21.5 are known as **class boundaries**. These class boundaries are obtained by

$$\begin{aligned} \text{Lower class boundary} &= \text{lower class limit} - \frac{d}{2}, \\ \text{Upper class boundary} &= \text{upper class limit} + \frac{d}{2}, \end{aligned}$$

where d = difference between any two consecutive classes. Here,

$$d = 1 \Longrightarrow \frac{d}{2} = 0.5.$$

(v) 2, 8, 14, 19 and 7 are called **class frequencies**.

(vi) The **class width** is the difference between the upper and lower class boundaries of a class interval. For example, the class width for the class interval 13–15 is

$$\text{Class width=15.5–12.5.=3.}$$

(vii) The **class mark** (or **midpoint**), x_m, of class interval is obtained by

$$x_m = \frac{\text{lower boundary + upper boundary}}{2}$$

or

$$x_m = \frac{\text{lower limit + upper limit}}{2}.$$

Example 1:

The following are the marks (out of 100) obtained by 50 students of MA131 in Semester II, 2003 Examination.

55	54	76	70	77	80	84	66	80	61
62	64	80	85	78	42	72	63	85	50
72	53	54	76	90	66	85	82	79	83
78	55	69	80	72	74	74	54	54	54
81	86	58	72	92	78	38	85	69	82

Construct a grouped frequency distribution. Use classes 30–39, 40–49, 50–59, etc.

Solution:

Mark interval	**Tally**	**No. of students** (f)
30–39	/	1
40–49	/	1
50–59	~~////~~ ~~////~~	10
60–69	~~////~~ ///	8
70–79	~~////~~ ~~////~~ ////	14
80–89	~~////~~ ~~////~~ ////	14
90–99	//	2
		50

Example 2:

The following is the distribution of the ages of new employees joined at a factory.

Age	**No. of employees**
20–29	7
30–39	21
40–49	4
50–59	2
60–69	1

(a) Obtain the class boundaries and class marks of the class intervals.

(b) What is the upper class limit of the class 30–39?

(c) What is the lower class boundary of the class 50–59?

(d) What is the class mark of the class 40–49?

Solution:

(a) The class boundaries and class marks are given in the following table:

Class Interval	**Class Boundary**	**Class Mark** (x_m)	**Frequency** (f)
20–29	19.5–29.5	24.5	7
30–39	29.5–39.5	34.5	21
40–49	39.5–49.5	44.5	4
50–59	49.5–59.5	54.5	2
60–69	59.5–69.5	64.5	1

(b) 39, (c) 49.5, (d) 44.5.

Example 3:

The following frequency distribution gives the lengths of 15 cucumbers.

Length (cm)	f
5–10	3
10–15	4
15–20	5
20–25	2
25–30	1

(a) What is the upper class limit of the class interval 15–20?

(b) What is the lower class boundary of the class 15–20?

(c) What is the class width of the class interval 15–20?

(d) What is the class mark of the class interval 15–20?

Solution: (a) 20, (b) 15, (c) 5, (d) 17.5.

Graphical Presentation

Graphs are used for presenting statistical data in an attractive way. They enable us to visualize the whole meaning of complex data at a single glance.

Pie charts:

A pie chart is an angular representation of a statistical data with several sub-divisions in a circle. The sub-divisions are made according to the percentages of several components. Pie charts are mostly used to represent the categorical (or qualitative) data.

Example 1:

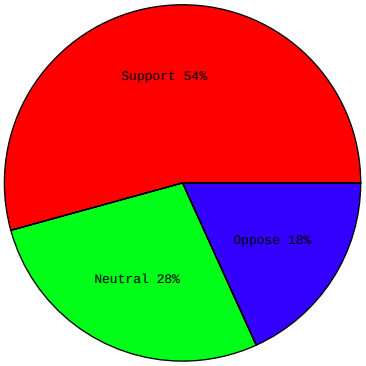
A campus press polled a sample of 280 undergraduate students in order to study student attitude toward a proposed change in the dormitory regulations. The following table shows the responses.

Response	**Frequency**
Support	152
Neutral	77
Oppose	51
Total	280

Draw a pie chart to show the students' responses.
Solution:

Calculation for Degrees and Percentages

Response	**Frequency**	**Degree**	**Percentage**
Support	152	$\frac{152}{280} \times 360^\circ = 195^\circ$	$\frac{152}{280} \times 100\% = 54\%$
Neutral	77	$\frac{77}{280} \times 360^\circ = 99^\circ$	$\frac{77}{280} \times 100\% = 28\%$
Oppose	51	$\frac{51}{280} \times 360^\circ = 66^\circ$	$\frac{51}{280} \times 100\% = 18\%$
Total	280	360°	100%



Example 2:

In an insurance company study of the causes of 1000 deaths, the following data were obtained.

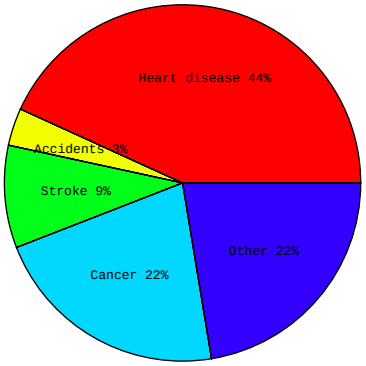
Cause of Death	**No. of Deaths**
Heart disease	432
Cancer	217
Stroke	93
Accidents	34
Other	224

Construct a pie graph to represent the data.

Solution:

Calculation for Degrees and Percentages

Cause of death	**No. of deaths**	**Degree**	**Percentage**
Hear diease	432	$\frac{432}{1000} \times 360^\circ = 156^\circ$	$\frac{432}{1000} \times 100\% = 44\%$
Cancer	217	$\frac{217}{1000} \times 360^\circ = 78^\circ$	$\frac{217}{1000} \times 100\% = 22\%$
Stroke	93	$\frac{93}{1000} \times 360^\circ = 33^\circ$	$\frac{93}{1000} \times 100\% = 9\%$
Accidents	34	$\frac{34}{1000} \times 360^\circ = 12^\circ$	$\frac{34}{1000} \times 100\% = 3\%$
Other	224	$\frac{224}{1000} \times 360^\circ = 81^\circ$	$\frac{224}{1000} \times 100\% = 22\%$
Total	1000	360°	100%



Bar graphs:

Bar graphs are the most popular ways to represent data graphically. One can see such graphs in newspapers, journals and even on television to depict different characteristics of data. For example, population, per capita income, sales and profits of a company, etc.
A bar graph consists of a series of bars of equal width. These bars are placed on common base line with equal gap. A bar graph may be either vertical or horizontal. The lengths of bars are proportional to the magnitudes that they represent.

The following types of bar graphs are commonly used:

- Simple bar graph
- Multiple bar graph
- Sub-divided bar graph

Simple bar graph:

This graph is used to show a single characteristic of different groups, categories or years.

Example 1:

In a study of 100 women, the numbers shown below indicate the major reason why each woman surveyed worked outside her home.

Reason	No. of Women
A. To support self/family	62
B. For extra money	18
C. For something different to do	12
D. Other	8

Construct a bar graph for the data.

Solution: