2. *Frequency Distribution and Graph*

2.1 Tabular Presentation:

There are two types of statistical tables:

i. **Simple table**: It is the table which distributed its data according to one characteristics or one phenomenon and may be summarized in two columns. The first columns of this table contains all values of the variable (qualitative or quantitative) and the second column indicate the corresponding frequency of occurrence.

Table 2: 100 students are distributed according to their weight.

Weight	No. of students
60-62	5
63-65	15
66-68	45
69-71	27
72-74	8
Total	100

ii. **Compound table**: It is the table which distributed the data according to two or more characteristics or phenomenon.

Table 3: 100 patients are distributed according to the weights and heights in compound table.

height \ weight	51-60	61-70	71-80	Total
121-140	20	6	4	30
141-160	2	40	10	52
161-180	2	6	10	18
Total	24	52	24	100

2.1 Some Important notes for data :

i. **Ungrouped data**: data that has not been placed in any categories and no aggregation/summarization has taken placed on the data then it is known as ungrouped data. Ungrouped data is also known as raw data. Ungrouped data is just in the form of number list.
For example, the marks scored obtained by 20 students in an examination are as follows

25, 37, 8, 40, 45, 29, 42, 12, 25, 16, 20, 36, 30, 33, 24, 24, 11, 35, 30, 45.

ii. **Grouped data**: Data that has been organized into groups (into a frequency distribution). If you see a table similar to the one below, you will know that you are dealing with grouped data.

Weight	No. of students
60-62	5
63-65	15
66-68	45
69-71	27
72-74	8
Total	100

2.1.1 Organizing Data

1- *Ordered Array*: Arranging data values from smallest to largest.

Example: Here are the ages of 10 people:

27 , 25 , 18 , 30 , 35 , 40 , 13 , 30 , 9 , 22

Ordered Array :

9 13 18 22 25 27 30 30 35 40

2-Frequency Distribution : In tabular presentation , we group the data into different classed and determining the number of observation that fall in each class, such arrangement in tabular form is called frequency distribution. Or , Organizing data into a table or a graph by putting them into non-overlapping classes (intervals) or categories.

Example:

The blood type of 20 patients released the following:

A, A , B ,O, AB , A , B , O , A, B , AB , A , A , A B, AB , B , A , AB , O.

Blood type	Frequency
A	8
B	5
AB	4
O	3

Example:

The numbers of medical tests that 17 patients did are shown below:

1,2, 0, 2, 1 ,3 , 2 , 2 , 1 , 3 , 4 , 2 , 1 , 2 , 2 , 1

Make a table to organize these data .

Number of tests	*F*
0	*1*
1	*6*
2	*7*
3	*2*
4	*1*

For Continuous Data:

i. **The range (R):** is the difference between the smallest and the largest observation in the data set.

ii. **Class**: it is a group which had divided to its variables values and every class takes the type of range from variables values (frequency) and sometimes we call it (Class interval = Class size), where:

$$\text{Class interval} = \frac{\text{Highest value} - \text{Lowest value}}{\text{Number of Classes}}$$

iii. **Class limit**: every class has two limits, lower limit and upper limit.

iv. **Class mark or midpoint**: it is the mid of range between limits and it can be found the following:

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

v. **No. of classes**: we can find the number of the classes by using the following:

$$\text{No. of classes} = 1 + 3.332\log(n) \ldots\ldots \text{Sturges method.}$$

Then we can calculated the class width (w) by the following:

$$\text{w} = \frac{\text{range(R)}}{\text{No. of classes}} = \frac{\text{Highest value} - \text{Lowest value}}{\text{No. of classes}}$$

vi. **Class Boundaries** : are class intervals but they always carried out to one more decimal place than the record observation so to ensures no that observation can fall precisely on the class boundary.

Note: The number of class should be not less than 5 and not more than 20 classes, because the first data lose some information and the second loss the purpose of grouped data in frequency table.

Example: Group the following raw data into ten classes, where the width number is 29.

Solution: The first step is to identify the width and lowest number

$$\text{Lowest number} = 1 \Longrightarrow \text{w} = \frac{29 - 1}{10} = 2.8$$

Example :

The amount of protein (in grams) for a variety of fast-food sandwiches is reported here. Construct a frequency distribution using 6 classes:

23	*30*	*20*	*27*	*44*	*26*	*35*	*20*	*29*	*29*	*25*	*15*	*18*	*27*	*19*	*22*	*12*	*26*	*34*
15	*27*	*35*	*26*	*43*	*35*	*14*	*24*	*12*	*23*	*31*	*40*	*35*	*38*	*57*	*22*	*42*	*24*	*21*	*27*	*33*

K = Number of classes

w= width of the class

$$\boldsymbol{w} = \frac{\textbf{\textit{The Range}}}{\textbf{\textit{number of classes}}} = \frac{\boldsymbol{X_L - X_S}}{\boldsymbol{K}}$$

$$\boldsymbol{w} = \frac{\text{Highest value} - \text{Lowest value}}{\text{No. of classes}} = \frac{57 - 12}{6} = 7.5 \cong 8$$

Class Limits	Frequency
12 – 19	7
20 – 27	17
28 – 35	10
36 – 43	4
44 – 51	1
52 – 59	1

The Boundaries and Class marks:

Class Boundaries	Class Marks OR Midpoint
11.5 – 19.5	15.5
19.5 – 27.5	23.5
27.5 – 35.5	31.5
35.5 – 43.5	39.5
43.5 – 51.5	47.5
51.5 - 59.5	55.5

Relative Frequency Distribution (RF)

The Relative Frequency for each class obtained by dividing the class frequency by the total frequency. That is

$$\text{Relative Frequency of class} = \frac{\text{Frequency of class}}{\text{sum of all frequency (total)}}$$

Percentage Frequency Distribution

The Percentage Frequency for each class obtained by multiplying the Relative Frequency by 100 % . That is

$$\text{Percentage Frequency of class} = \frac{\text{Frequency of class}}{\text{sum of all frequency (total)}} * 100\%$$

Class Limits	Relative Frequency RF	Percentage Relative Frequency
12 – 19	7/40 = 0.175	17.5 %
20 – 27	17/40 = 0.425	42.5 %
28 – 35	0.25	25 %
36 – 43	0.1	10 %
44 – 51	0.025	2.5 %
52 – 59	0.025	2.5 %

Cumulative Frequency Distribution (CF):

The total frequency of all values less than the upper C.B of a given class is Called the cumulative frequency up to and including that class and denoted.by CF. A Table listing cumulative Frequencies is Called cumulative Frequency distribution.

Classes	*Cumulative Frequency*
less than 11.5	*0*
less than 19.5	*7*
less than 27.5	*24*
less than 35.5	*34*
less than 43.5	*38*
less than 51.5	*39*
less than 59.5	*40*

Example: The following table is Frequency Distribution of ages of 100 patients.

Age (years)	No. of patients
21-25	5
26-30	18
31-35	42
36-40	27
41-45	8
Total	100

Find the Relative Frequency and Class mark.

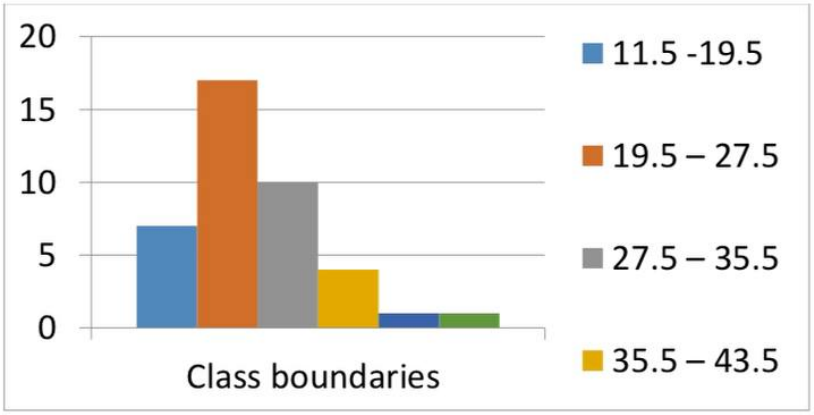
Solution

Age (years)	No. of patients	Relative Frequency	Class mark
21-25	5	0.05	23
26-30	18	0.18	28
31-35	42	0.42	33
36-40	27	0.27	28
41-45	8	0.08	43
Total	100	1	

2.2.2 Graphical Representation

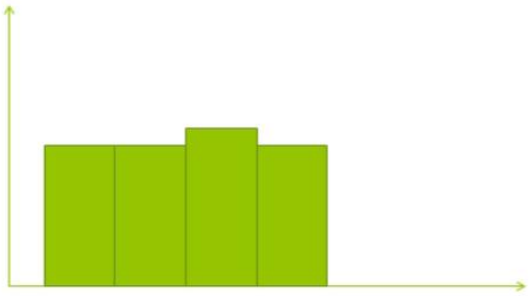
- **Histogram :** used to represent data Graphically

Example : Draw a histogram for the last example



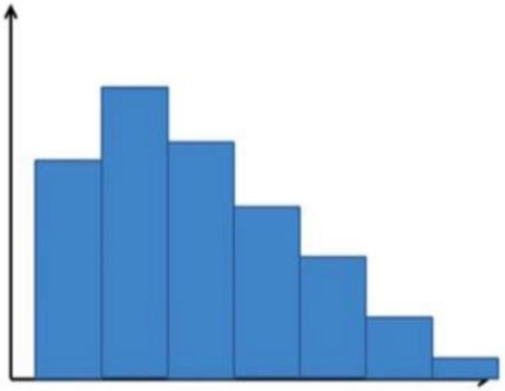
Distribution Shapes

1- Uniform Distribution

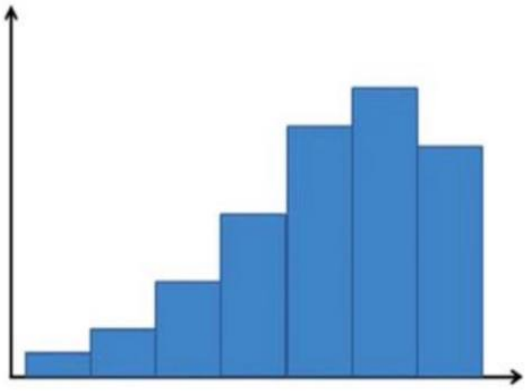


2- Bell shaped distribution

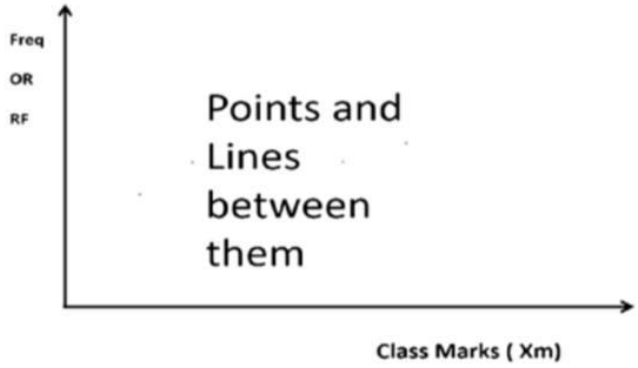
3- Right(positively) skewed distribution



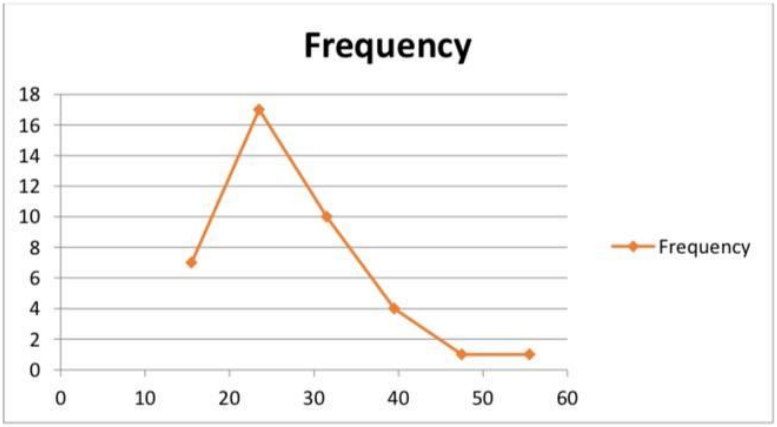
4- Left(negatively) skewed distribution



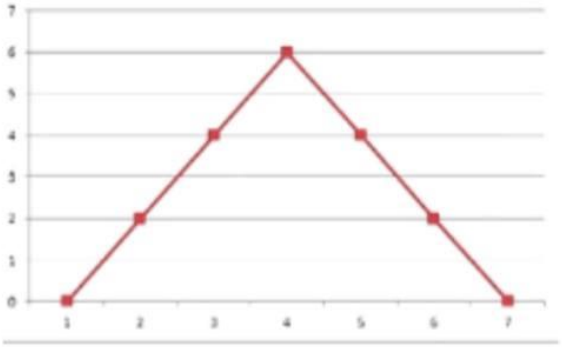
- ***The polygon (Closed Line graph)***

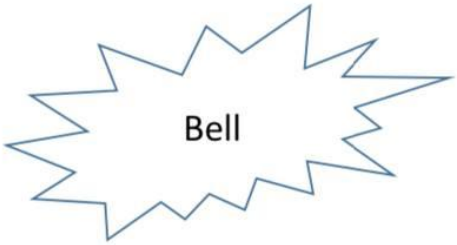


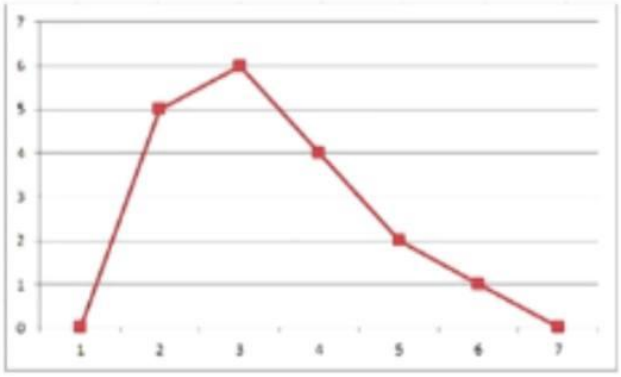
Example : Draw a Polygon for the last example



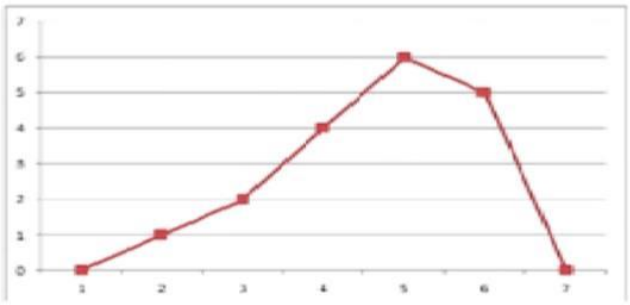
Distribution Shapes



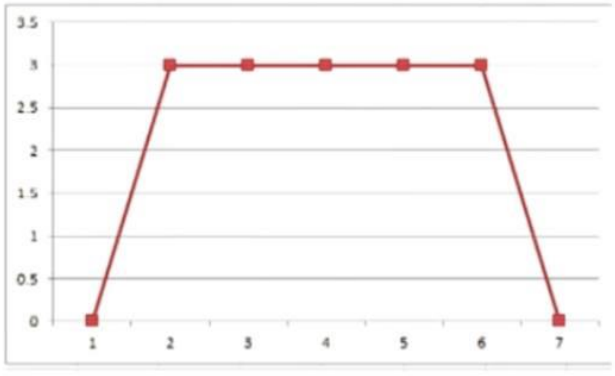




Right



Left



Uniform

H.W

1- The following data shows the number of hours 45 hospital patients slept following the administration of a certain anesthetic.

7	10	12	4	8	7	3	8	5	12	11	3	8
1	1	13	10	4	4	5	5	8	7	7	3	2
3	8	13	1	7	17	3	4	5	5	3	1	17
10	4	7	7	11	8							

Find the Frequency table, Relative Frequency and Class mark.

2- Fill the space in table

Percentage %	No. of tables	Relative Frequency	Class mark
Less than 90	10		
90-96	15		
97-101	3		
102-106	12		
107-110	5		
Over 110	5		
Total	50		