CHAPTER -3

3. METHODS OF DATA PRESNTATION

Having collected and edited the data, the next important step is to organize it. That is to present it in a readily comprehensible condensed form that aids in order to draw inferences from it. It is also necessary that the like be separated from the unlike ones.

The presentation of data is broadly classified in to the following two categories:

- Tabular presentation
- Diagrammatic and Graphic presentation.

The process of arranging data in to classes or categories according to similarities technically is called *classification*.

Classification is a preliminary and it prepares the ground for proper presentation of data.

Definitions:

- Raw data: recorded information in its original collected form, whether it be counts or measurements, is referred to as raw data.
- Frequency: is the number of values in a specific class of the distribution.
- Frequency distribution: is the organization of raw data in table form using classes and frequencies.

There are three basic types of frequency distributions

- Categorical frequency distribution
- Ungrouped frequency distribution
- Grouped frequency distribution

There are specific procedures for constructing each type.

1)   Categorical frequency Distribution:

Used for data that can be place in specific categories such as nominal, or ordinal. e.g. marital status.
Example: a social worker collected the following data on marital status for 25
        persons. (M=married, S=single, W=widowed, D=divorced)

```
M    S    D    W    D
S    S    M    M    M
W    D    S    M    M
W    D    D    S    S
S    W    W    D    D
```

Solution:

Since the data are categorical, discrete classes can be used. There are four types of marital status M, S, D, and W. These types will be used as class for the distribution. We follow procedure to construct the frequency distribution.

Step 1: Make a table as shown.

Class	Tally	Frequency	Percent
(1)	(2)	(3)	(4)
M			
S			
D			
W			

Step 2: Tally the data and place the result in column (2).

Step 3: Count the tally and place the result in column (3).

Step 4: Find the percentages of values in each class by using;

$$\% = \frac{f}{n} * 100$$ Where f= frequency of the class, n=total number of value.

Percentages are not normally a part of frequency distribution but they can be added since they are used in certain types diagrammatic such as pie charts.

Step 5: Find the total for column (3) and (4).

Combing all the steps one can construct the following frequency distribution.

Class	Tally	Frequency	Percent
(1)	(2)	(3)	(4)
M	////	5	20
S	//// //	7	28
D	//// //	7	28
W	////	6	24

2) Ungrouped frequency Distribution:

-Is a table of all the potential raw score values that could possible occur in the data along with the number of times each actually occurred.

-Is often constructed for small set or data on discrete variable.

Constructing ungrouped frequency distribution:

- First find the smallest and largest raw score in the collected data.
- Arrange the data in order of magnitude and count the frequency.
- To facilitate counting one may include a column of tallies.

Example:

The following data represent the mark of 20 students.

80	76	90	85	80
70	60	62	70	85
65	60	63	74	75
76	70	70	80	85

Construct a frequency distribution, which is ungrouped.

Solution:

Step 1: Find the range, Range=Max-Min=90-60=30.
Step 2: Make a table as shown
Step 3: Tally the data.
Step 4: Compute the frequency.

Mark	Tally	Frequency
60	//	2
62	/	1
63	/	1
65	/	1
70	////	4
74	/	1
75	//	2
76	/	1
80	///	3
85	///	3
90	/	1

Each individual value is presented separately, that is why it is named ungrouped frequency distribution.

3) Grouped frequency Distribution:

-When the range of the data is large, the data must be grouped in to classes that are more than one unit in width.

Definitions:

- **Grouped Frequency Distribution:** a frequency distribution when several numbers are grouped in one class.
- **Class limits:** Separates one class in a grouped frequency distribution from another. The limits could actually appear in the data and have gaps between the upper limits of one class and lower limit of the next.
- **Units of measurement (U):** the distance between two possible consecutive measures. It is usually taken as 1, 0.1, 0.01, 0.001, -----.
- **Class boundaries:** Separates one class in a grouped frequency distribution from another. The boundaries have one more decimal places than the row data and therefore do not appear in the data. There is no gap between the upper boundary of one class and lower boundary of the next class. The lower class boundary is found by subtracting U/2 from the corresponding lower class limit and the upper class boundary is found by adding U/2 to the corresponding upper class limit.
- **Class width**: the difference between the upper and lower class boundaries of any class. It is also the difference between the lower limits of any two consecutive classes or the difference between any two consecutive class marks.
- **Class mark (Mid points):** it is the average of the lower and upper class limits or the average of upper and lower class boundary.
- **Cumulative frequency:** is the number of observations less than/more than or equal to a specific value.
- **Cumulative frequency above:** it is the total frequency of all values greater than or equal to the lower class boundary of a given class.
- **Cumulative frequency blow:** it is the total frequency of all values less than or equal to the upper class boundary of a given class.
- **Cumulative Frequency Distribution (CFD):** it is the tabular arrangement of class interval together with their corresponding cumulative frequencies. It can be more than or less than type, depending on the type of cumulative frequency used.
- **Relative frequency (rf):** it is the frequency divided by the total frequency.
- **Relative cumulative frequency (rcf):** it is the cumulative frequency divided by the total frequency.

Guidelines for classes

1. There should be between 5 and 20 classes.
2. The classes must be mutually exclusive. This means that no data value can fall into two different classes

3. The classes must be all inclusive or exhaustive. This means that all data values must be included.
4. The classes must be continuous. There are no gaps in a frequency distribution.
5. The classes must be equal in width. The exception here is the first or last class. It is possible to have an "below ..." or "... and above" class. This is often used with ages.

Steps for constructing Grouped frequency Distribution

1. Find the largest and smallest values
2. Compute the Range(R) = Maximum - Minimum
3. Select the number of classes desired, usually between 5 and 20 or use Sturges rule $k = 1 + 3.32 \log n$ where *k is number of classes desired and n is total number of observation.*
4. Find the class width by dividing the range by the number of classes and rounding up, not off. $w = \dfrac{R}{k}$.
5. Pick a suitable starting point less than or equal to the minimum value. The starting point is called the lower limit of the first class. Continue to add the class width to this lower limit to get the rest of the lower limits.
6. To find the upper limit of the first class, subtract U from the lower limit of the second class. Then continue to add the class width to this upper limit to find the rest of the upper limits.
7. Find the boundaries by subtracting U/2 units from the lower limits and adding U/2 units from the upper limits. The boundaries are also half-way between the upper limit of one class and the lower limit of the next class. !may not be necessary to find the boundaries.
8. Tally the data.
9. Find the frequencies.
10. Find the cumulative frequencies. Depending on what you're trying to accomplish, it may not be necessary to find the cumulative frequencies.
11. If necessary, find the relative frequencies and/or relative cumulative frequencies

Example*:

Construct a frequency distribution for the following data.

11	29	6	33	14	31	22	27	19	20
18	17	22	38	23	21	26	34	39	27

Solutions:

Step 1: Find the highest and the lowest value H=39, L=6

Step 2: Find the range; R=H-L=39-6=33

Step 3: Select the number of classes desired using Sturges formula;

$$k = 1 + 3.32 \log n = 1 + 3.32 \log (20) = 5.32 = 6 \text{(rounding up)}$$

Step 4: Find the class width; w=R/k=33/6=5.5=6 (rounding up)

Step 5: Select the starting point, let it be the minimum observation.

- 6, 12, 18, 24, 30, 36 are the lower class limits.

Step 6: Find the upper class limit; e.g. the first upper class=12-U=12-1=11

- 11, 17, 23, 29, 35, 41 are the upper class limits.

So combining step 5 and step 6, one can construct the following classes.

Class limits
6 – 11
12 – 17
18 – 23
24 – 29
30 – 35
36 – 41

Step 7: Find the class boundaries;

E.g. for class 1 Lower class boundary=6-U/2=5.5

Upper class boundary =11+U/2=11.5

- Then continue adding w on both boundaries to obtain the rest boundaries. By doing so one can obtain the following classes.

Class boundary
5.5 – 11.5
11.5 – 17.5
17.5 – 23.5
23.5 – 29.5
29.5 – 35.5
35.5 – 41.5

Step 8: tally the data.

Step 9: Write the numeric values for the tallies in the frequency column.

Step 10: Find cumulative frequency.

Step 11: Find relative frequency or/and relative cumulative frequency.

The complete frequency distribution follows:

Class limit	Class boundary	Class Mark	Tally	Freq.	Cf (less than type)	Cf (more than type)	rf.	rcf (less than type
6 – 11	5.5 – 11.5	8.5	//	2	2	20	0.10	0.10
12 – 17	11.5 – 17.5	14.5	//	2	4	18	0.10	0.20
18 – 23	17.5 – 23.5	20.5	/////	7	11	16	0.35	0.55
24 – 29	23.5 – 29.5	26.5	////	4	15	9	0.20	0.75
30 – 35	29.5 – 35.5	32.5	///	3	18	5	0.15	0.90
36 – 41	35.5 – 41.5	38.5	//	2	20	2	0.10	1.00

Diagrammatic and Graphic presentation of data.

These are techniques for presenting data in visual displays using geometric and pictures.

Importance:

- They have greater attraction.
- They facilitate comparison.
- They are easily understandable.

-Diagrams are appropriate for presenting discrete data.

-The three most commonly used diagrammatic presentation for discrete as well as qualitative data are:

- Pie charts
- Pictogram
- Bar charts

Pie chart

- A pie chart is a circle that is divided in to sections or wedges according to the percentage of frequencies in each category of the distribution. The angle of the sector is obtained using:

$$Angle\ of\ sector = \frac{Value\ of\ the\ part}{the\ whole\ quantity} * 360$$

Example: Draw a suitable diagram to represent the following population in a town.

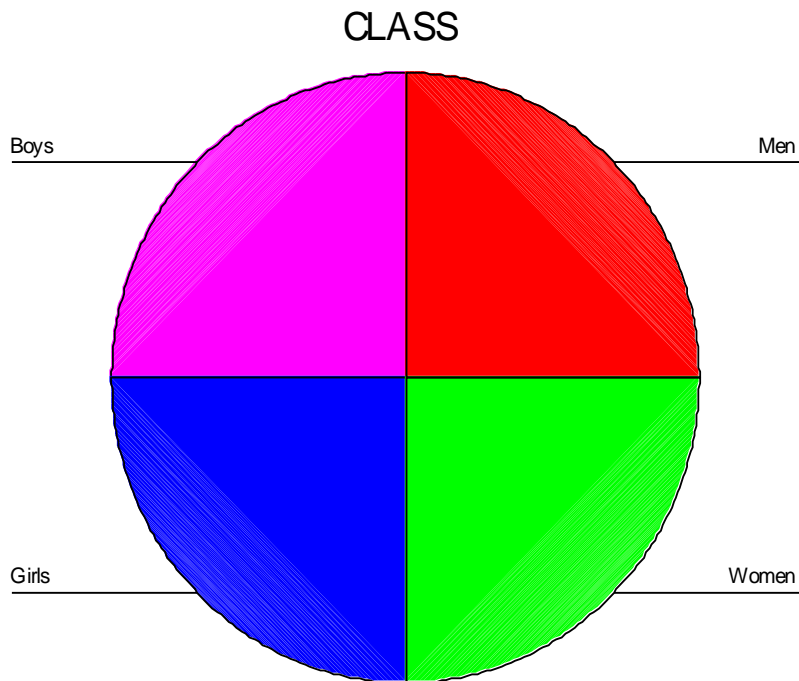| Men	Women	Girls	Boys
2500	2000	4000	1500

Solutions:

Step 1: Find the percentage.

Step 2: Find the number of degrees for each class.

Step 3: Using a protractor and compass, graph each section and write its name corresponding percentage.

Class	Frequency	Percent	Degree
Men	2500	25	90
Women	2000	20	72
Girls	4000	40	144
Boys	1500	15	54

<div align="center">

CLASS

</div>

Boys ─────────   ───────── Men

Girls ─────────   ───────── Women

Pictogram

-In these diagram, we represent data by means of some picture symbols. We decide abut a suitable picture to represent a definite number of units in which the variable is measured.

Example: draw a pictogram to represent the following population of a town.

Year	1989	1990	1991	1992
Population	2000	3000	5000	7000

Bar Charts:
- A set of bars (thick lines or narrow rectangles) representing some magnitude over time space.
- They are useful for comparing aggregate over time space.
- Bars can be drawn either vertically or horizontally.
- There are different types of bar charts. The most common being :
  - Simple bar chart
  - Deviation o0r two way bar chart
  - Broken bar chart
  - Component or sub divided bar chart.
  - Multiple bar charts.

Simple Bar Chart

-Are used to display data on one variable.

-They are thick lines (narrow rectangles) having the same breadth. The magnitude of a quantity is represented by the height /length of the bar.

Example: The following data represent sale by product, 1957- 1959 of a given company for three products A, B, C.

| Product	Sales($) In 1957	Sales($) In 1958	Sales($) In 1959
A	12	14	18
B	24	21	18
C	24	35	54

Solutions:
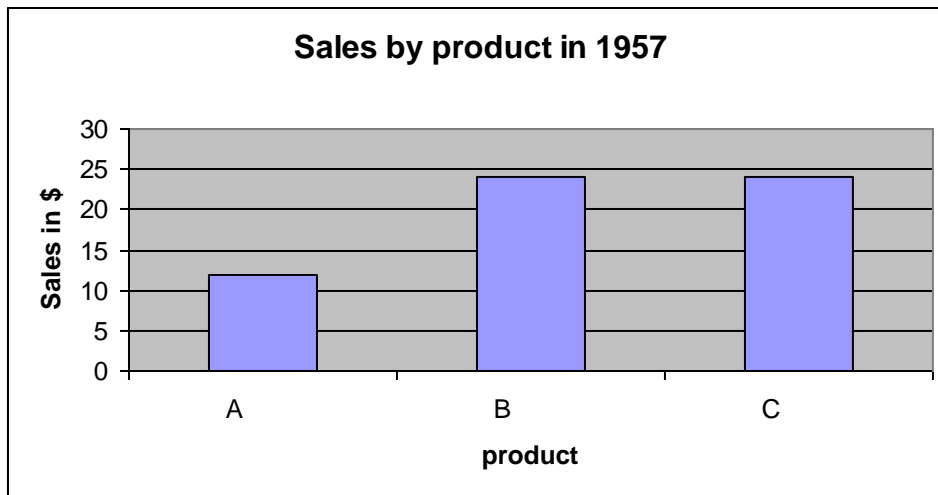
Component Bar chart

-When there is a desire to show how a total (or aggregate) is divided in to its component parts, we use component bar chart.

-The bars represent total value of a variable with each total broken in to its component parts and different colours or designs are used for identifications

Example:

Draw a component bar chart to represent the sales by product from 1957 to 1959.
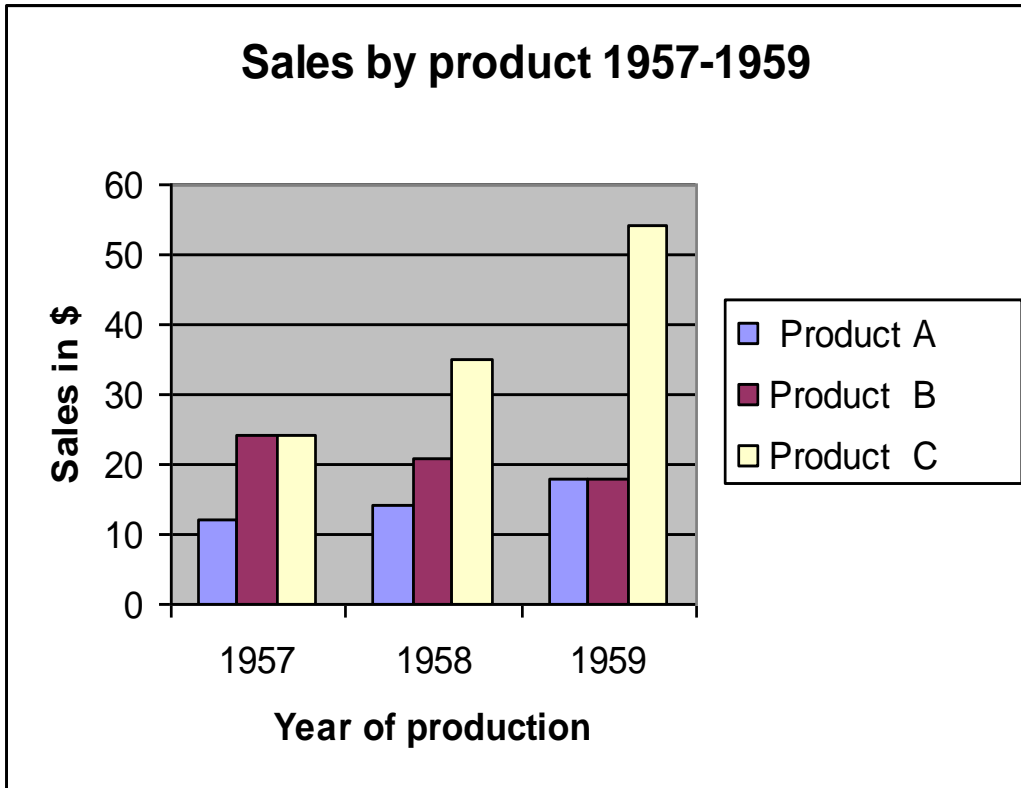
Solutions:

SALES BY PRODUCT 1957-1959
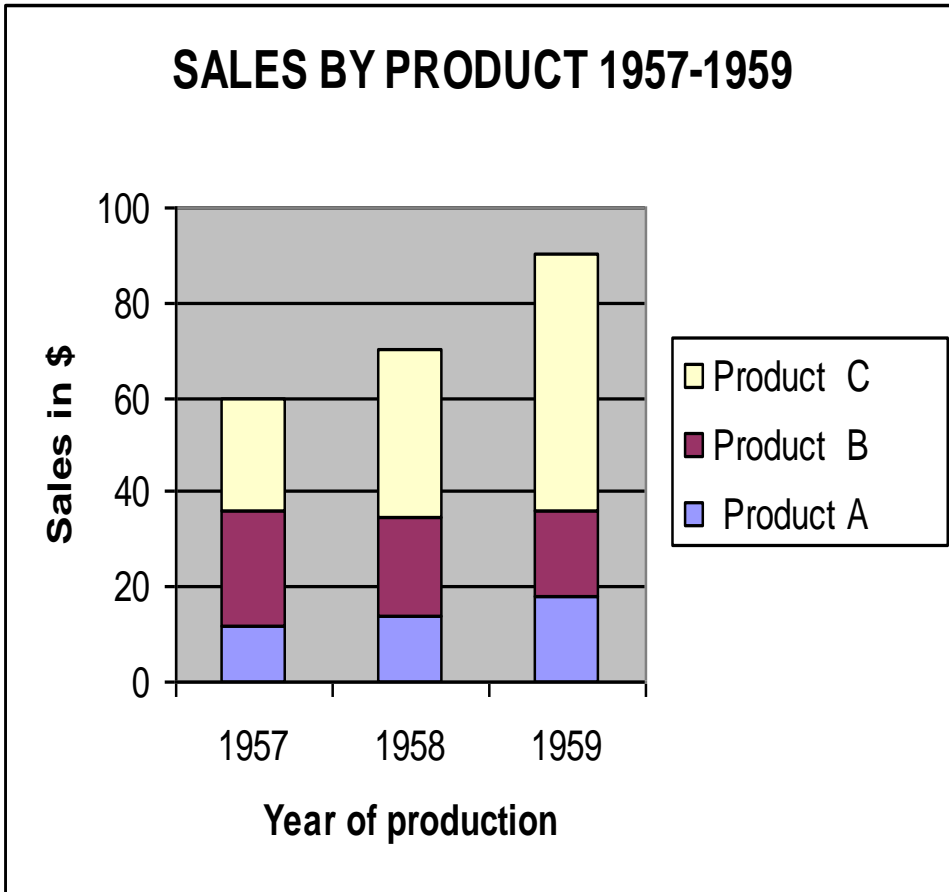
Multiple Bar charts
- These are used to display data on more than one variable.
- They are used for comparing different variables at the same time.

Example:

Draw a component bar chart to represent the sales by product from 1957 to 1959.

Solutions:

Sales by product 1957-1959

Graphical Presentation of data

- The histogram, frequency polygon and cumulative frequency graph or ogive are most commonly applied graphical representation for continuous data.

Procedures for constructing statistical graphs:

- Draw and label the X and Y axes.
- Choose a suitable scale for the frequencies or cumulative frequencies and label it on the Y axes.
- Represent the class boundaries for the histogram or ogive or the mid points for the frequency polygon on the X axes.
- Plot the points.
- Draw the bars or lines to connect the points.

Histogram

A graph which displays the data by using vertical bars of various heights to represent frequencies. Class boundaries are placed along the horizontal axes. Class marks and class limits are some times used as quantity on the X axes.

Example: Construct a histogram to represent the previous data (example *).
  Frequency Polygon:

- A line graph. The frequency is placed along the vertical axis and classes mid points are placed along the horizontal axis. It is customer to the next higher and lower class interval with corresponding frequency of zero, this is to make it a complete polygon.

Example: Draw a frequency polygon for the above data (example *).

Solutions:

Class Mid points
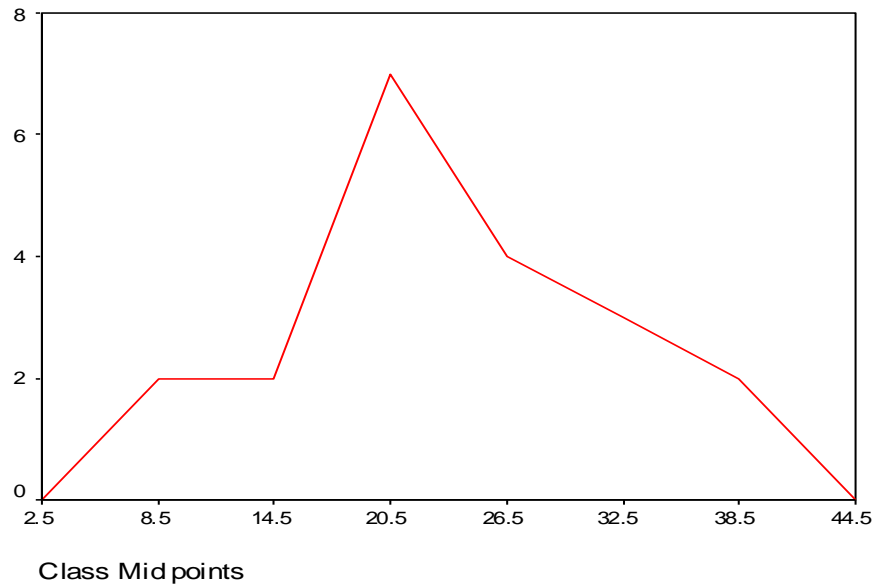
Ogive (cumulative frequency polygon)
- A graph showing the cumulative frequency (less than or more than type) plotted against upper or lower class boundaries respectively. That is class boundaries are plotted along the horizontal axis and the corresponding cumulative frequencies are plotted along the vertical axis. The points are joined by a free hand curve.

Example: Draw an ogive curve(less than type) for the above data.
(Example *)