
بسم الله الرحمن الرحيم

PROBABILITY & STATISTICS
MS-301
SOFTWARE ENGINEERING

Lecture # 2

(Presentation of Data)

Prepared by Dr. Asma Zaffar

Graphical methods for presenting data

2.1 Introduction

We have looked at ways of collecting data and then collating them into tables. Frequency tables are useful methods of presenting data; they do, however, have their limitations. With large amounts of data graphical presentation methods are often clearer to understand. Here, we look at methods for producing graphical representations of data of the types we have seen previously.

2.3 Bar Charts

Bar charts are a commonly–used and clear way of presenting categorical data or any ungrouped discrete frequency observations.
For example, recall the example on students' modes of transport:

Student	Mode	Student	Mode	Student	Mode
1	Car	11	Walk	21	Walk
2	Walk	12	Walk	22	Metro
3	Car	13	Metro	23	Car
4	Walk	14	Bus	24	Car
5	Bus	15	Train	25	Car
6	Metro	16	Bike	26	Bus
7	Car	17	Bus	27	Car
8	Bike	18	Bike	28	Walk
9	Walk	19	Bike	29	Car
10	Car	20	Metro	30	Car

The first logical step is to put these into a frequency table, giving

Mode	Frequency
Car	10
Walk	7
Bike	4
Bus	4
Metro	4
Train	1
Total	30

We can then present this information as a bar chart, by following the five step process shown below:

1. First decide what goes on each axis of the chart. By convention the variable being measured goes on the horizontal (x–axis) and the frequency goes on the vertical (y–axis).

2. Next decide on a numeric scale for the frequency axis. This axis represents the frequency in each category by its height. It must start at zero and include the largest frequency. It is common to extend the axis slightly above the largest value so you are not drawing to the edge of the graph.

3. Having decided on a range for the frequency axis we need to decide on a suitable number scale to label this axis. This should have sensible values, for example, $0, 1, 2, \ldots$, or $0, 10, 20 \ldots$, or other such values as make sense given the data.

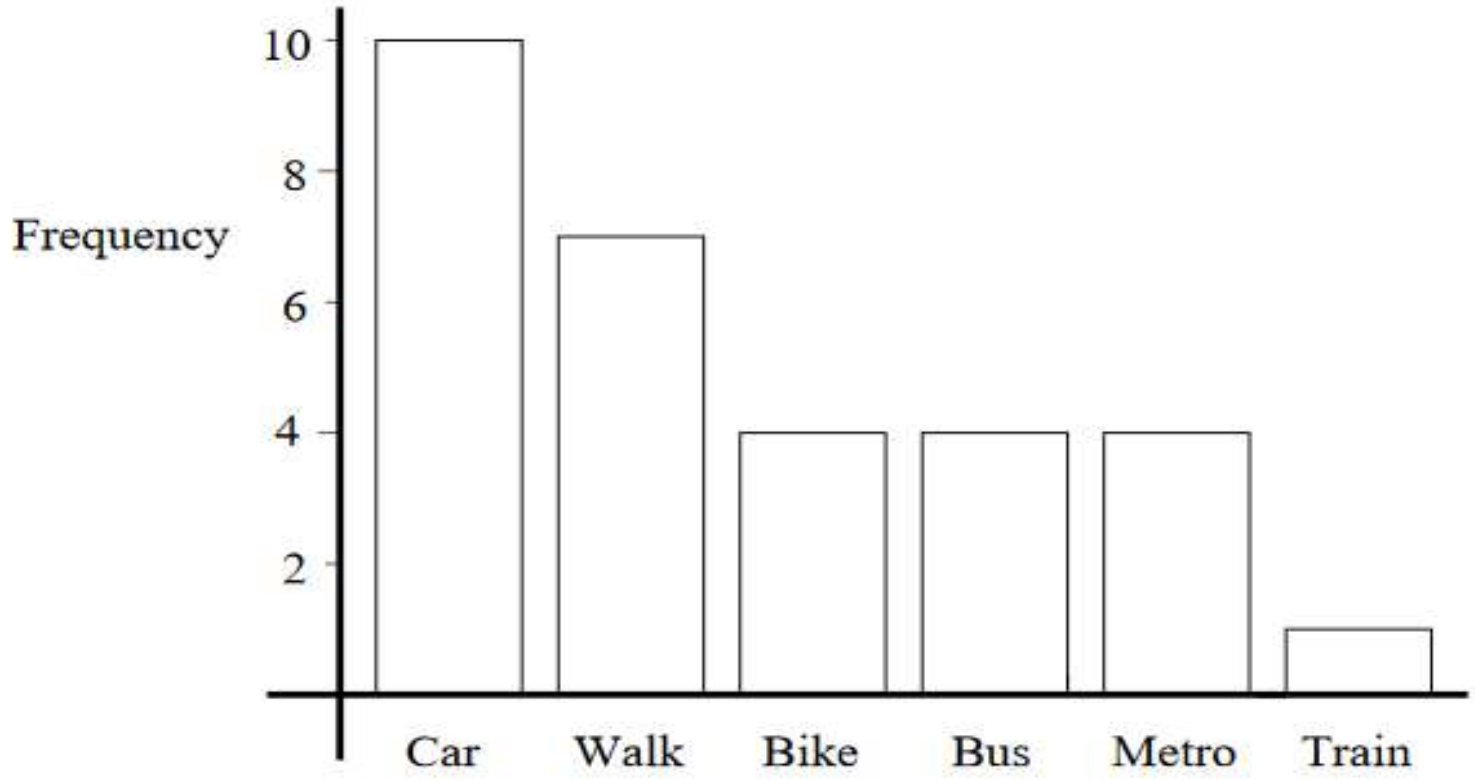
4. Draw the axes and label them appropriately.

5. Draw a bar for each category. When drawing the bars it is essential to ensure the following:

   – the width of each bar is the same;

   – the bars are separated from each other by equally sized gaps.

This gives the following bar chart:



This bar chart clearly shows that the most popular mode of transport is the car and that the metro, bus and cycling are all equally popular (in our small sample). Bar charts provide a simple method of quickly spotting simple patterns of popularity within a discrete data set.

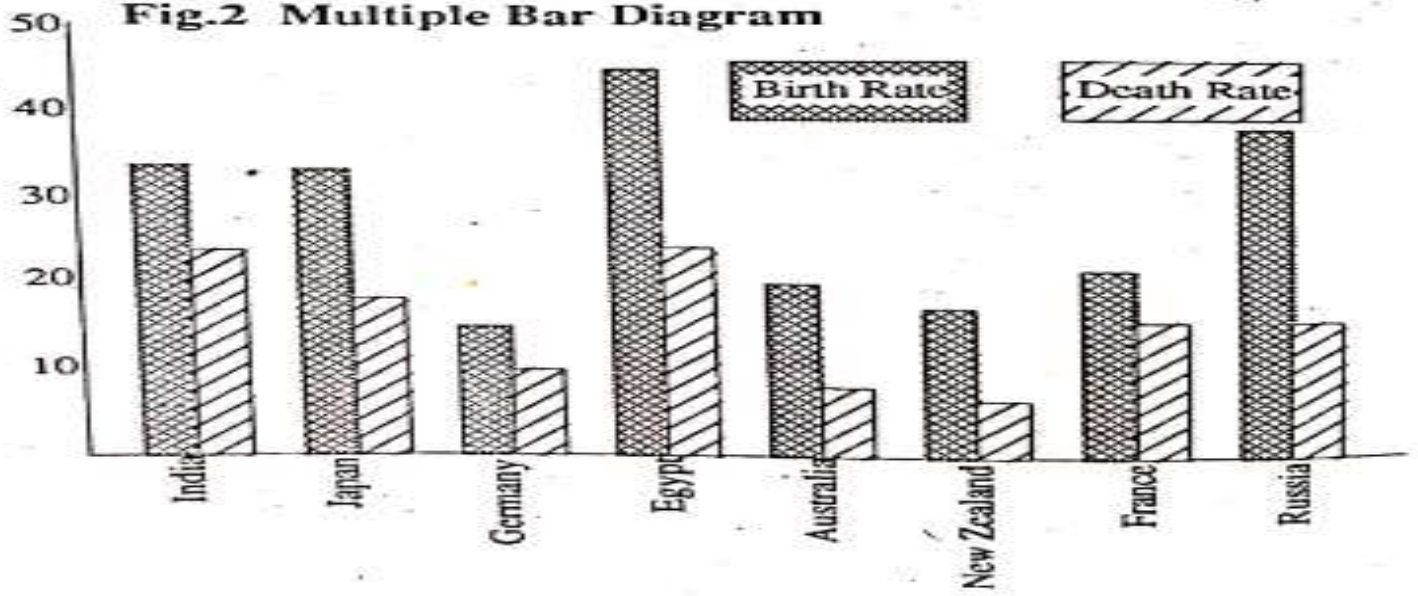
EXAMPLE 8 **The following table gives the birth and death rates per thousand of a few countries. Represent this data by multiple bar chart.**

Country	Birth Rate	Death Rate
India	33	24
Japan	32	19
Germany	16	10
Egypt	44	24
Australia	20	9
New Zealand	18	8
France	21	16
Russia	38	16

(Gujranwala, B.I.S.E; 1987)

The required diagram is given in Figure 2.

Fig.2 Multiple Bar Diagram



Year	Wheat	Barley	Oats	Total
1971	34	18	27	79
1972	43	14	24	81
1973	43	16	27	86
1974	45	13	34	92

The required diagram is given in Figure 3.

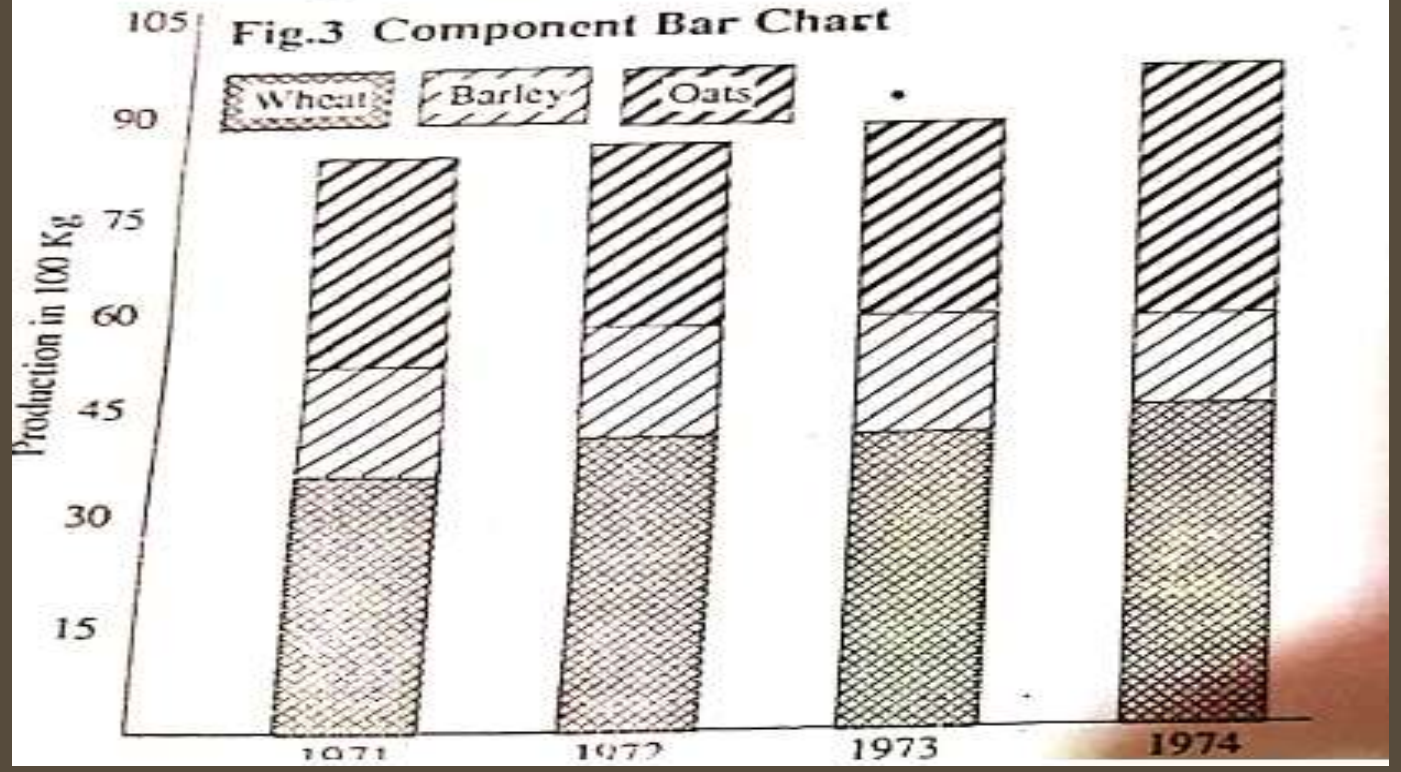


Fig.3 Component Bar Chart

1.2.1 Pie Chart

A **pie chart** is a circular chart divided into sectors, illustrating relative magnitudes in frequencies or percents. In a pie chart, the area is proportional to the quantity it represents.

Example 1.4. *As the nation debates strategies for delivering health insurance, let's look at the sources of funds and the types of expenditures.*

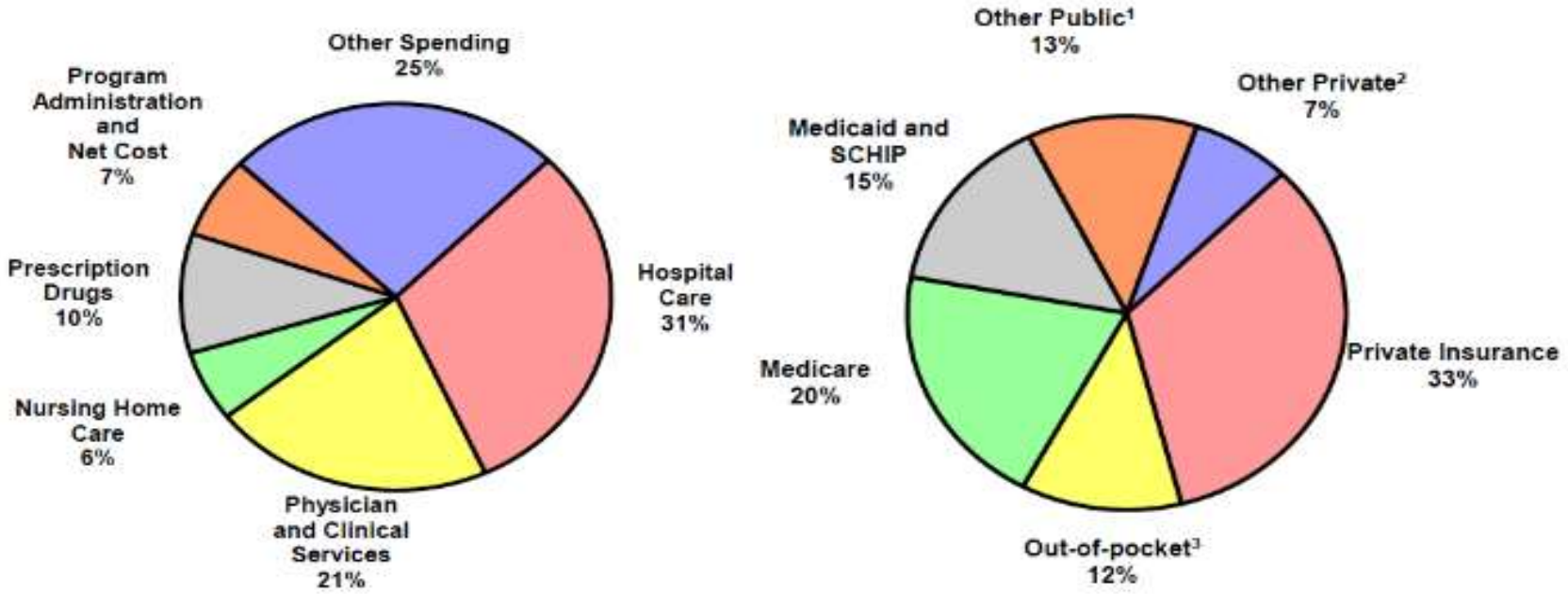


Figure 1.1: 2008 United States health care (a) expenditures (b) income sources, Source: Centers for Medicare and Medicaid Services, Office of the Actuary, National Health Statistics Group

1.4 Histograms and the Empirical Cumulative Distribution Function

Histograms are a common visual representation of a quantitative variable. Histograms summarize the data using rectangles to display either frequencies or proportions as normalized frequencies. In making a histogram, we

- Divide the range of data into bins of equal width (usually, but not always).
- Count the number of observations in each class.
- Draw the histogram rectangles representing frequencies or percents by *area*.

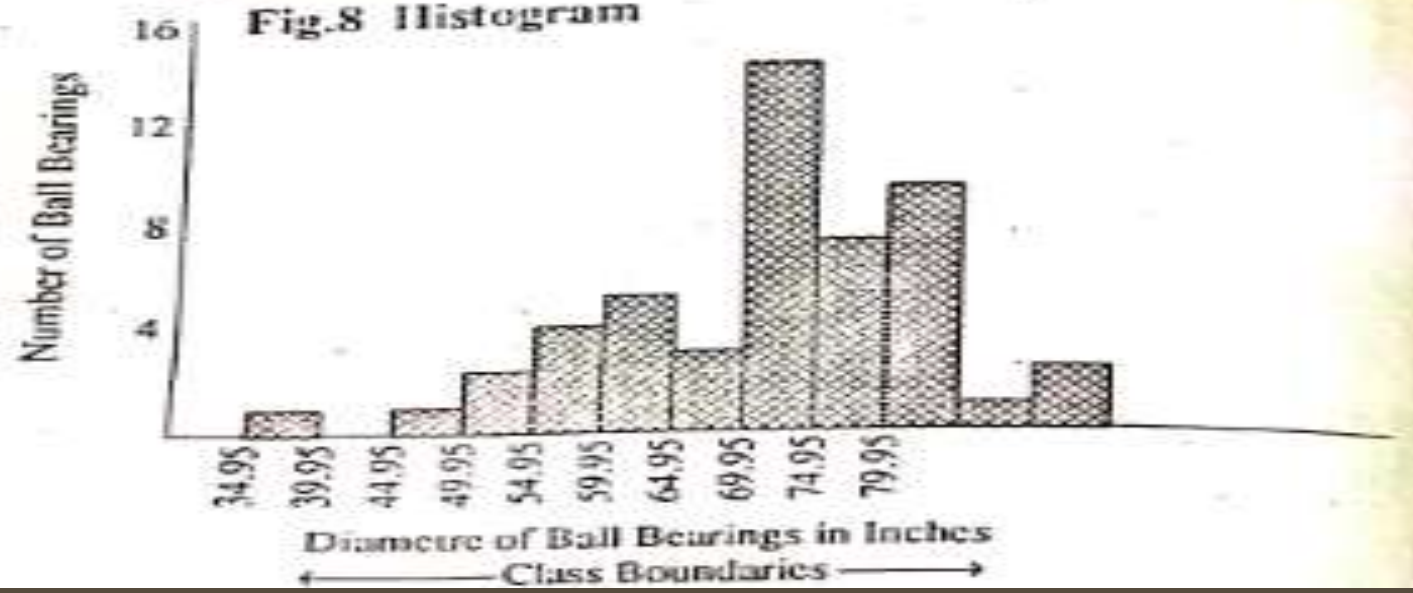

Histogram of age
Frequency
0
5
10
15
40
45
50
55
60
65
70
age

Frequency distribution of Ball - bearings by Diameter

Diameter of Ball - bearings (Class limits)	No. of Ball bearings (f)	Class Boundaries (C.B)
35 – 39.9	1	34.95 – 39.95
40 – 44.9	1	39.95 – 44.95
45 – 49.9	2	44.95 – 49.95
50 – 54.9	4	49.95 – 54.95
55 – 59.9	5	54.95 – 59.95
60 – 64.9	3	59.95 – 64.95
65 – 69.9	14	64.95 – 69.95
70 – 74.9	7	69.95 – 74.95
75 – 79.9	9	74.95 – 79.95
80 – 84.9	1	79.95 – 84.95
85 – 89.9	3	84.95 – 89.95
90 – 94.9		89.95 – 94.95

The desired histogram is given in figure 8.

Fig.8 Histogram



2.11.2.2 FREQUENCY POLYGON AND FREQEUNCY CURVE

Frequency Polygon and Frequency curve both are graphs of frequency distribution but less commonly used as compared to histogram. Following steps are involved in the construction of frequency polygon and freqeuncy curve:

(i) Take mid points or class marks along x–axis and frequency along y–axis.

(ii) Place a dot against the each class mark with respect to its class freqeuncy.

(iii) Join the dots by means of straight line (foot rûle) to get freqeuncy polygon or join the dots by smoothed line to get frequency curve.

(iv) Extend frequency polygon at both ends to touch x–axis at next class mid–point with zero frequencies.

Another way of obtaining frequency polygon or curve is by joining the mid–point of the tops of rectangles in the histogram.

EXAMPLE 15 Draw freqeuncy polygon and frequency curve representing the following data:

Marks:	10 – 19,	20 – 29,	30 – 39,	40 – 49,	50 – 59
f:	5,	25,	40,	20,	10

SOLUTION Calcualte the mid–points of each class:

Marks	f	Mid – Points
10 – 19	5	14.5
20 – 29	25	24.5
30 – 39	40	34.5
40 – 49	20	44.5
50 – 59	10	54.5

The desired frequency polygon and frequency curve are given in figure 10–A and 10–B.

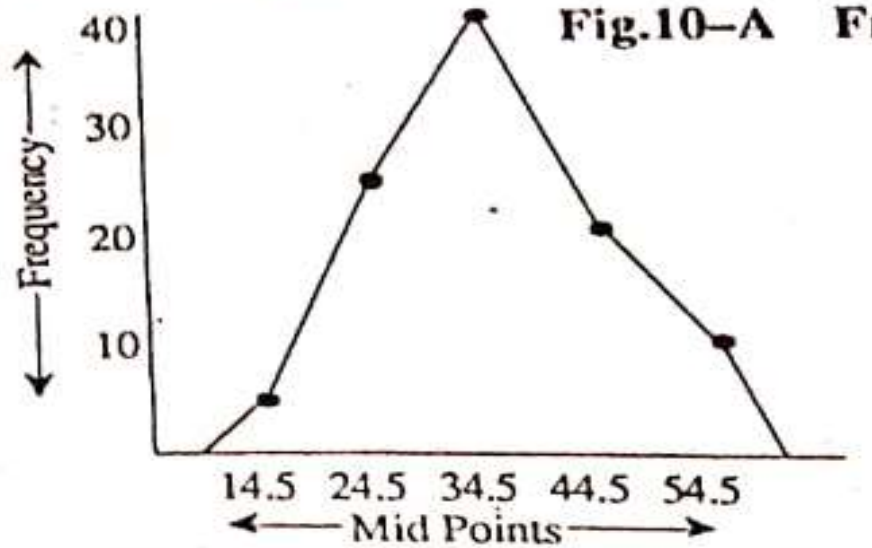


Fig.10–A Frequency Ploygon

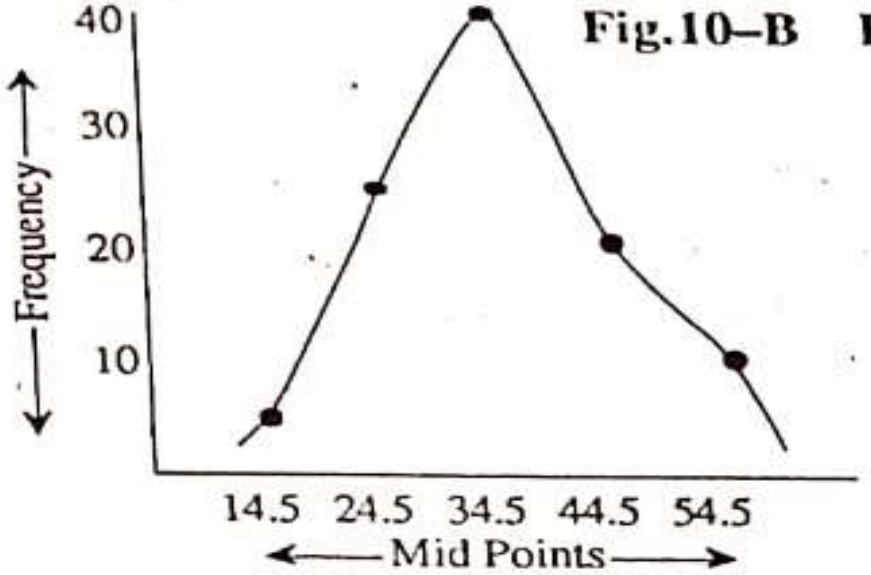


Fig.10–B Frequency Curve

CONSTRUCTION OF FREQUENCY DISTRIBUTION

2.8.1 HOW TO MAKE A DISCRETE FREQUENCY DISTRIBUTION

To prepare a discrete frequency distribution, one has to go through the following steps:

(i) Find the largest and smallest value of the given data.

(ii) Place all possible values of the variable from the smallest to largest in a column with the title of the variable under consideration.

(iii) Put a bar (vertical line) opposite the particular value to which it relates. To facilitate counting prepared blocks of five bars, and some space should be left in between each of the blocks. The column prepared in this way is titled as tally marks or tally bars.

(iv) Finally, count the number of bars corresponding to each value of the variable and place it in the column entitled "frequency".

EXAMPLE 2 **The following are the number of flowers on different branches of a tree. Make a discrete frequency distribution:**

2, 4, 6, 1, 3, 3, 5, 7, 8, 6, 4, 7, 4, 4, 2, 1, 3, 6, 4, 2, 5, 7, 9, 1; 2, 10, 1, 8, 9, 2, 3, 1, 2, 3, 4, 4, 4, 6, 6, 5, 5, 4, 5, 8, 5, 4, 3, 3, 2, 5, 0, 5, 9, 9, 8, 10, 0, 4, 10, 1, 1, 2, 2, 1, 8, 6, 9, 10.

(Gujranwala, B.I.S.E; 1989)

SOLUTION The given variable is "number of flowers on different branches" which is a discrete variable.

The largest and smallest values of the given data are '0' and "10" respectively.

The desired discrete frequency distribution is given in table 9.

TALLY TABLE

No. of Flowers	Tally Bars	No. of Branches
0	II	2
1	𝍸 III	8
2	𝍸 IIII	9
3	𝍸 II	7
4	𝍸 𝍸 I	11
5	𝍸 III	8
6	𝍸 II	7
7	III	3
8	IIII	4
9	𝍸	5
10	IIII	4
Total	—	68

TABLE 9
Desired Frequency Distribution

No. of Flowers	No. of Branches
0	2
1	8
2	9
3	7
4	11
5	8
6	7
7	3
8	4
9	5
10	4
Total	68

EXAMPLE 4 The data given below shows the diameter in inches of ball – bearing manufactured by a company.

73.1, 78.3, 56.6, 78.5, 84.9, 74.7, 63.1, 73.8, 74.7, 70.8, 84.2, 84.1, 80.1, 86.4, 66.9, 64.2, 61.3, 58.4, 67.8, 78.0, 64.5, 84.0, 71.2, 56.3, 74.2, 72.4, 83.4, 94.2, 58.2, 70.4, 84.6, 77.2, 93.7, 82.1, 39.3, 74.6, 84.7, 72.1, 72.0, 54.1, 48.1, 78.1, 64.2, 94.0, 78.3, 66.0, 77.5, 71.5, 70.4, 51.0.

Prepare a freqeuncy distribution using the following groups 35 – 39.9, 40 – 44.9.....

SOLUTION Largest value = 94.2

Smallest Value = 39.3

As the pattern of classes is given so we do not need to find class – interval and number of possible classes.

The desired frequency distribution is given in table 11(b).

TABLE 11 – A
Tally Table

Diameter of Ball-bearing (in inches)	Tally Bar	Frequency
35–39.9	I	1
40–44.9	–	–
45–49.9	I	1
50–54.9	II	2
55–59.9	IIII	4
60–64.9	IIII	5
65–69.9	III	3
70–74.9	IIII IIII IIII	14
75–79.9	IIII II	7
80–84.9	IIII IIII	9
85–89.9	I	1
90–94.9	III	3
Total	—	50

TABLE 11 – B
Frequency Distribution of students Height

Diameter of Ball-bearing (in inches)	No. of Ball-bearing
35 – 39.9	1
40 – 44.9	–
45 – 44.9	1
50 – 54.9	2
55 – 59.9	4
60 – 64.9	5
65 – 69.9	3
70 – 74.9	14
75 – 79.9	7
80 – 84.9	9
85 – 89.8	1
90 – 84.9	3
Total	50

Questions?