

Description of Module

Subject Name	Environmental Sciences
Paper Name	Statistical Applications in Environmental Sciences
Module Name/Title	Diagrammatic and Graphical Representation of Data II
Module Id	EVS/SAES-XIV/5
Pre-requisites	Module 4 Diagrammatic and Graphical Representation of Data II
Objectives	Types of different diagrammatic representation of data
Keywords	Histogram, frequency polygon, ogive curves



Pathshala
पाठशाला
A Gateway to All Post Graduate Courses

Module 5: Diagrammatic and Graphical Representation of Data II

- **Learning Objectives**
- **Introduction.**
- **Graphical presentation of frequency distribution**
- **Stem and Leaf Plot**
- **Pie Chart**
- **Summary**
- **Suggested Readings**

1. Learning Objectives

In this module, a complete explanation about different types of representation of data will be discussed. This module helps to learn different methods of diagrammatic and graphical presentation and their properties. Through this module, one can learn about which method to use under what conditions. Questions with answers are included to give an in-depth knowledge of the topic.

2. Introduction

In this module, one can get a depth knowledge of different types of diagram like pie chart and graphical representations of the data like histogram, frequency polygon etc. This module is continuation of the previous module. Hence it is desired that reader must go through the previous module before further reading. We have already discussed about the frequency distribution and its different forms in the previous module. In this module, graphical representation of the frequency distribution and its types will be discussed. Other diagrammatic and graphical methods and their properties are also discussed. This module will help to develop an understanding of different methods of graphical and diagrammatic presentation of the data. In the following section, graphical presentation of frequency distributions are discussed in detail.

3. Graphical presentation of frequency distribution

Graphs which consider the frequencies for the presentation of the data are called frequency graphs. These graphs are used to describe the characteristic feature of the frequency data. The methods for the preparation of the frequency table is already discussed in the previous module. The frequency graphs usually depends on the type of data whether it is discrete or continuous. Discrete data are usually represented in the form of bar diagrams or discontinuous curves while the continuous data are represented by means of curves.

While constructing the frequency graphs, usually the value of the variables are taken on the x-axis and the corresponding frequencies are represented on y-axis. The graphs that are used for this purpose are:-

- (i) Histogram
- (ii) Frequency Polygon
- (iii) Ogives

(i) **Histogram:** It is among the most commonly used method for the diagrammatic representation of a grouped frequency distribution. On one axis it consists of bars with height proportional to class frequencies and the width of the bar is defined through the class boundaries presented on the horizontal scale. In general, same width is considered for all the bars but it is not essential as it depends on the class width. In histogram, there is difference in terms of height only due to the class frequencies. The area of each bar is calculated by multiplying the width of the class and the frequency.

Sometimes people are confused about bar diagram and histogram. The main difference between the bar diagram and histogram is that there is no difference between the bars in histogram while in bar diagram there is a gap between the bars. Also the bars in the bar diagram are of equal width but in histogram it depends on the width of classes and may be unequal due to unequal class width. Also it should be remembered that histogram can only be constructed for open ended class width if magnitude of first open class is same as next class and the magnitude of the last class is same as preceding class.

The following example helps to get an idea about the construction of the histogram. The data is given in the following table:

Class Marks	10-15	15-20	20-25	25-30	30-35	35-40	40-45	45-50	50-55	55 more
Frequency	5	6	11	12	15	13	10	6	7	4

Table 1

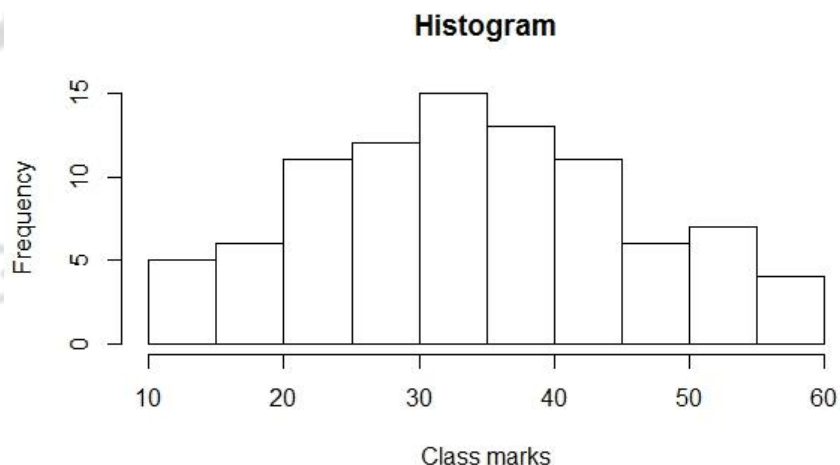


Figure 1

From the above figure, one can observe that class marks are shown on the x-axis and the corresponding frequency is available on y-axis. The height of the bar is based on the frequency value. Also from the Figure 1, one can notice that there is no gap between the bars because values in the tables are shown in

intervals and there is no gap in the upper limit of one class and the lower limit of the succeeding class. So to represent the continuous data in an exact manner bars are represented in the figure without any gap. On the other hand bar diagram is basically used to show the frequency of individual observations. Hence there is gap in the bars due to difference between the observations to maintain the difference. This is the major difference between histogram and the bar diagram. The following example will help to understand the differences between the two in detail.

Observations	10	15	20	25	30	35	40	45	50	55
Frequency	5	6	11	12	15	13	10	6	7	4

Table 2

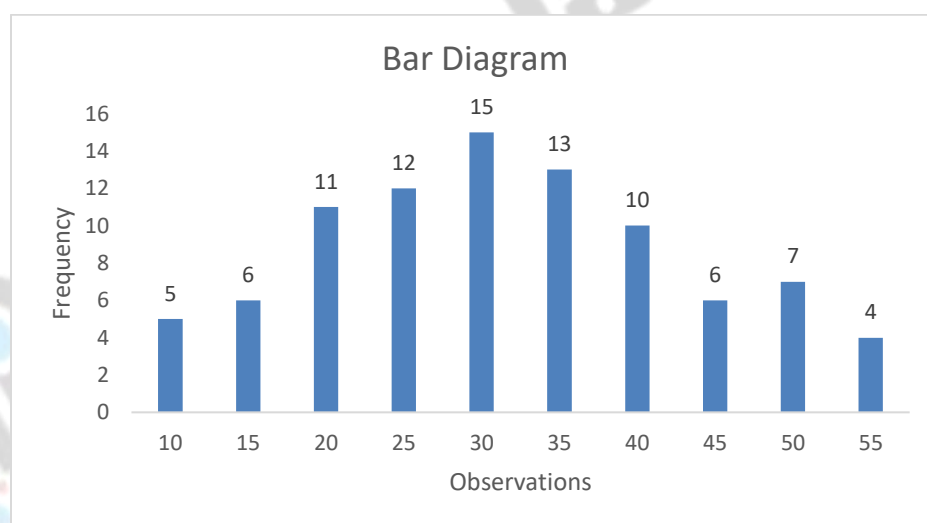


Figure 2

From the above bar diagram, one can observe that there is gap in the bars that justify our discussion on histogram and bar diagram.

Hence, one can understand the difference between histogram and bar chart by keeping in mind the above examples.

Now as we are familiar with histogram, another most commonly used method for the graphical presentation of the data is frequency polygon.

- (ii) **Frequency Polygon:** Frequency Polygon is considered as another method, other than histogram for graphical representation of the data. It is generally plotted when all the classes have same width. It is obtained by joining the mid value of the class interval corresponding to the class frequency. The end points of the polygon are joined to the x-axis that is basically the mid-point of classes at end of the

frequency distribution. Hence frequency polygon is used in the situation of same class width and covered similar area as histogram.

Frequency polygon helps in understanding the skewness property of the data. One can detect the mode value from frequency polygon by drawing a perpendicular from the highest point on the curve to the x-axis. The frequency polygon of various distributions can be plotted on the same axis for comparison purpose but this is not possible in the case of histogram. In Histogram, if we have to compare two histogram of different data sets then we require two separate graphs. Also frequency polygon is a continuous curve and it has other benefits like determination of the slope, rough estimates of a particular value on the x-axis, at what rate observations are changing. Another important feature of the frequency polygon is that it can be drawn without plotting the histogram only when the mid-points of the classes are known.

In the following graph shown in Figure 3, a frequency polygon is constructed from Table 1 data. In the graph, one can observe that it is a curve that is represented with the joining of mid points of class interval and corresponding class frequency on the y-axis.

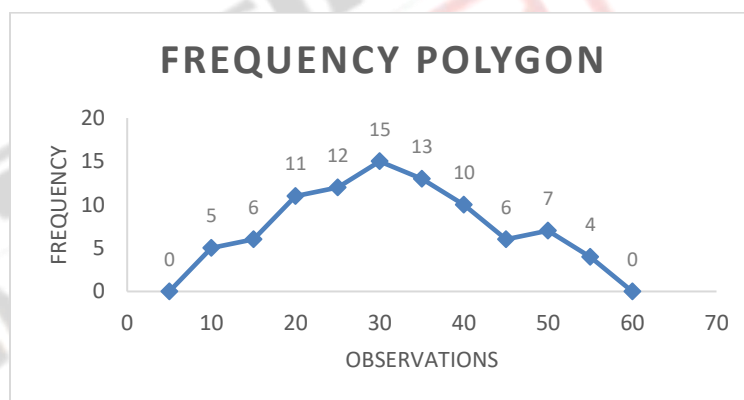


Figure 3

Frequency Curve

Frequency curve is the term that is basically used for the presentation of the probability distribution of continuous variable. As continuous variable takes values in small interval and the range of these variables approaches to infinite. So if we decrease the width of class intervals and at the same time increase the total frequency of the data that approaches to infinite then the shape of the histogram and the frequency polygon is approximate close to the shape of the frequency curve. In that case, the horizontal axis will represent the range of the variable and the vertical axis will show the frequency of a values in an interval. If it is feasible to present the relative frequency of the dataset on the vertical axis then it shows the percentage of particular class interval. This method of representing the data will help you to understand the concept of probability distribution function in future where we have probability values on the vertical axis that is derived from a mathematical function. On the other hand, the range of continuous variable is shown on horizontal axis. As the variable is continuous so the probability of a particular point is zero. So probability are evaluated in terms of intervals. So one can correlate the

probability distribution curve and the frequency curve from the concept. The topic of probability distribution function will be discussed in the further modules.

As we already discussed about the frequency curve, there are basically five types of frequency curves in the literature. These are

- (a) Symmetrical bell shaped (b) Uniform (c) U-shaped (d) J shaped and (e) asymmetric (Positive or negative skew)

- (a) **Symmetrical bell shaped:** In this type of curve, the class frequency values keep on increasing steadily and after attaining a maximum value it keep on decreasing in the same pace as while increasing. It can be checked whether the curve is symmetric if we fold the curve from the centre then the two half the frequency curve must coincide. This is the reason that such type of curve is called symmetrical curve that has same area under curve on both side of the middle part or centre value. Another feature of symmetric curve is that it has single peakness at the middle and lessen slowly on both end.

SHAPES of Frequency Curves

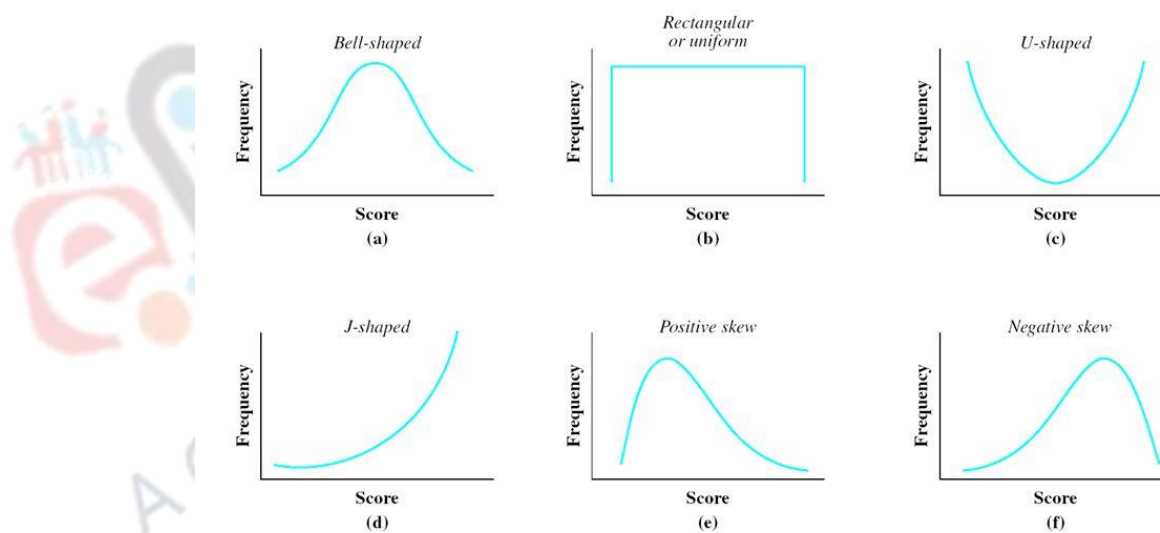


Figure 4

Link for resource

<http://slideplayer.com/slide/4929425/>

- (b) **Uniform Curve:** This type of curve will only occur if all the observations have same frequency. So it means when occurrence of each observation in the dataset is same then such type of curve appear. For example, in Figure 4 (b) figure represent the score of students of class. If these is a great possibility that all students score equal marks in the class test or examination. In such the uniform frequency curve will appear.
- (c) **U-Shaped:** In this type of frequency curve, the highest value of the frequency occurs at the end or extreme point and the frequency keep on decreasing as the till the central part of the data and keep on increasing at the same pace of decreasing before reaching the middle part. This curve is just the opposite of the symmetric curve in shape. In this curve the most of the values occur at the extreme point and the less values come from intermediate point.
- (d) **J-Shaped:** In this curve, the value of the variable attains the maximum frequency at one of its extreme points only. Hence it is different from U-shaped curve where maximum value attained at both extreme values. In a J-shaped curve, the values has less frequency in the lower class interval and then frequency keep on increasing steadily as the value of variable increases. Hence at the extreme point the distribution attains its maximum value.
- (e) **Asymmetric:** This type of frequency curve has the highest frequency not in the central part of the data but on the either side of the central value. If the curve has longer tail toward the right side then the distribution is said to be positively skewed. If the curve has longer tail towards the left side then the distribution is said to be negative skewed.
- (iii) **Ogive:** Ogive curve is constructed from the cumulative values of frequency i.e. cumulative frequency values are plotted against the x-axis. There are two types of ogive curves **less than** or **more than**. **Less than** ogive curve begin from the first interval and keep on increasing upwards corresponding to the next interval as the frequency are cumulated that means frequency values of previous frequency are added in next frequency values. Hence it keeps on increasing and at the last interval it shows the total frequency of data plotted against the last interval. Similarly, **more than** ogive curve begin from the total cumulative value from the first interval and keep on decreasing at next interval by subtracting the previous interval frequency from the total frequency value. If we repeat this process till the last interval then one can observe that the curve seem to be like a elongated S and is sometimes calls a double curve with one portion being concave and the other being convex. Also ogive curve can be constructed for unequal width of classes in the frequency distribution.
- It is also possible to construct **more than** or **less than** ogive curve on the same graph and one can observe that if both curve are drawn on the same graph, these curve intersect at a point. X-axis value of that point is the median value of the frequency distribution. Hence ogive curve will help to get an idea about characteristics of the data as it is only the diagrammatic representation

of the cumulative frequency distribution. There are some of the uses of the ogive which are as follows:

- Ogive curve is used to find out the median of the frequency distribution. It is also used to find out quartiles, deciles and percentiles etc.
- It is used to observe the cumulative frequency of a particular class interval of more than and less than type.
- From the shape of the curve one can observe the number of observations which are expected to lie between two given values.

It should be noted that it may be possible that one may not be able to estimate the correct value from the ogive curve. Hence, one should be careful while using ogives.

An example will help to understand the concept in depth. The cumulative frequency distribution (less than type) of data used in previous examples are shown below:

Class Marks	10-15	15-20	20-25	25-30	30-35	35-40	40-45	45-50	50-55	55 more
Frequency	5	6	11	12	15	13	10	6	7	4
Cumulative Frequency (less than)	5	11	22	34	49	62	72	78	85	89

Table 3

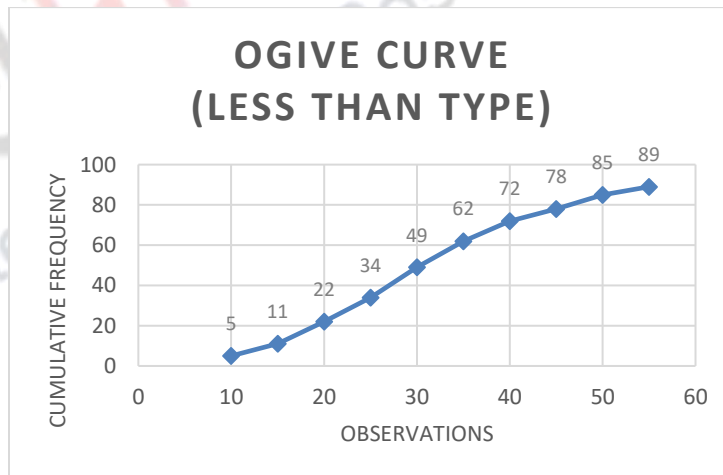


Figure 5

In Figure 5, ogive curve represents the cumulative frequency of less than type and it is of increasing type. We use the same data for graphical representation of the more than type of ogive type to see the difference of two type of ogive curve.

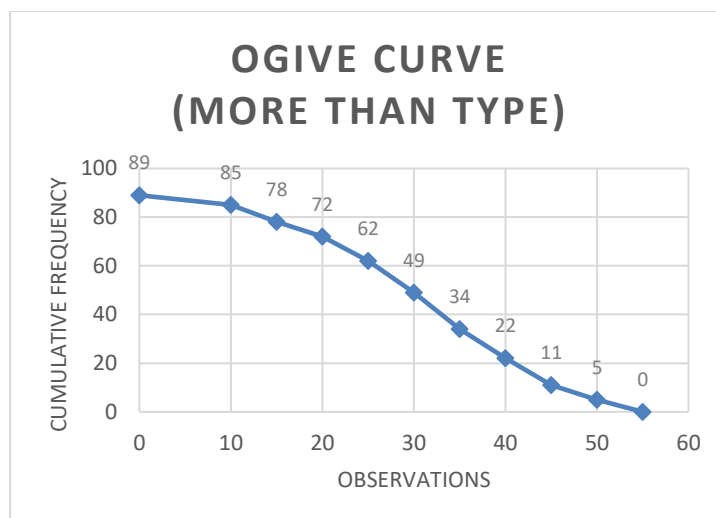


Figure 6

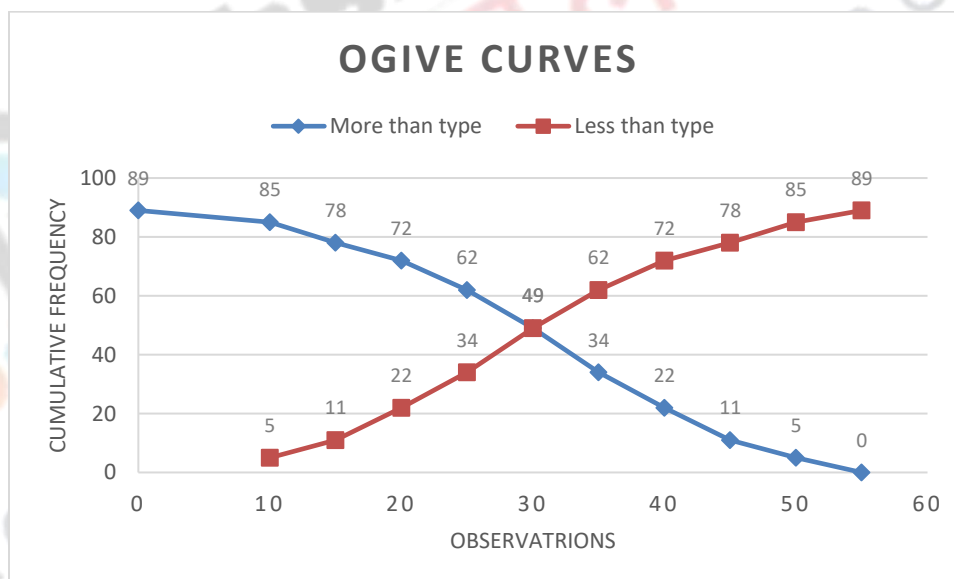


Figure 7

In Figure 7, ogive curve of two types are plotted graphically. From intersection point of these curve draw a perpendicular to the cumulative frequency polygon, horizontal scale i.e. these two curves intersect at point value 49 at y-axis and approx. 30 at x-axis. This approximate value is considered as the median value of the data. In the next section, we will discuss Stem and leaf display that is considered as another method for the representation of the data.

4. Stem and Leaf plot

It is considered as an alternative to the histogram. Although it gives a visual representation similar to the histogram but it does not lose the details of the individual data points in the grouping. For example, the data of expenditure on calls in a week of 30 persons are given below:

11,22,22,25,27,28,29,34,35,36,37,37,38,39,39,39,43,46,49,49,49,49,52,56,58,64,65,74,82,91

Stem	Leaf
1	1
2	225789
3	456778999
4	369999
5	268
6	45
7	4
8	2
9	1

Table 4

The stems on the left represent the 10 units and the leaves on the right represent the units of ones. So the individual data points can be represented in the diagram. From the diagram, one can observe that it seems like a histogram but the difference here is that the values are used in place of bars. Hence the major drawback of this diagram is that when we have large amount of then it is difficult to make stem and leaf diagram.

In the next section, we will discuss about the diagrammatical representation of the data through the pie chart.

5. Pie-Chart

When the total values can be represented by a big circle and the various components by sectors cut inside it. This type of diagram is known as pie diagram and it is also used to show the percentage breakdown. The total is represented by 360 degree angle. It can be divided into a number of small angles whose degree is according to the values of the categories in the data. There are some steps for the construction of pie diagram. These are

- The category or items values are expressed either in percentage or degree. For example, number of males/females in the data are expressed either in percentage or in degrees.
- One can find the percentage of the items by dividing its value by the aggregate one and multiple each of them by 3.6.

Hence one can find two forms of pie chart, one is in percentage form and other in degree term.

For example, the following data represent the expenditure on entertainment and communication of household

Items	Expenditure in Rs.	Items	Expenditure
Cable	250	Telephone bill	250
Mobile bill	1000	Movie Ticket	1000
Internet bill	750	Total	3250

Table 5

Items	Expenditure	Percentage
Cable	250	7.69
Mobile	1000	30.76
Internet bill	750	23.07
Telephone bill	250	7.69
Movie Ticket	1000	30.76
Total	3250	

Table 6

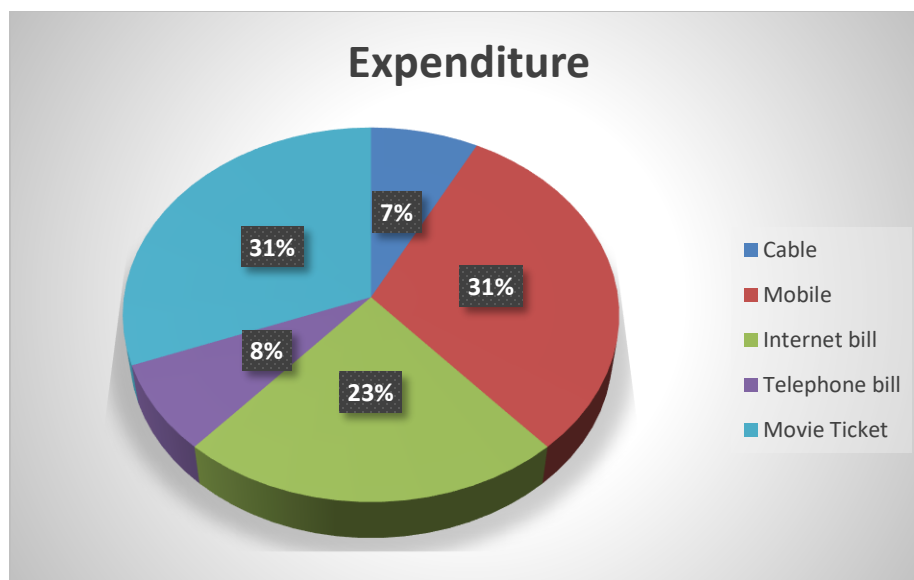


Figure 8

Hence one can easily understand the expenditure on items through percentages from the pie diagram. This is the reason that pie chart is mostly used to present the data as it is easily understandable even by a layman.

6. Summary

In the real world large type of diagrammatic methods are used for the presentation of the statistical data. Some of the types are like one- dimensional diagram, two-dimensional diagram, three-dimensional diagram and pictogram etc. In this module, we discussed some of its types like histogram, bar diagram, frequency polygon, frequency curve, ogives and pie charts. Differences between histogram and bar diagram are discussed. Importance of all these graphical and diagrammatical representation is discussed with examples. Frequency curve and its different types are also discussed. These diagrams and graphical representations are useful as they present the data in an attractive manner that appeal more to the mind of the spectators. These forms are more attractive, fascinating and impressive than the other methods. The best part of diagrammatic representation method is that even a layman can understand this without any previous knowledge of statistics. This is the reason that diagrams, pictures and graphs are used to give primary education to the kids.

7. Suggested Readings

Agresti, A. and B. Finlay, Statistical Methods for the Social Science, 3rd Edition, Prentice Hall, 1997.

Daniel, W. W. and C. L. Cross, C. L., Biostatistics: A Foundation for Analysis in the Health Sciences, 10th Edition, John Wiley & Sons, 2013.

Hogg, R. V., J. Mckean and A. Craig, Introduction to Mathematical Statistics, Macmillan Pub. Co. Inc., 1978.

Meyer, P. L., Introductory Probability and Statistical Applications, Oxford & IBH Pub, 1975.

Triola, M. F., Elementary Statistics, 13th Edition, Pearson, 2017.

Weiss, N. A., Introductory Statistics, 10th Edition, Pearson, 2017.

