

## **Graphical Representation of Data**

### **Meaning, Principles and Methods**

#### **Meaning of Graphic Representation of Data:**

Graphic representation is another way of analysing numerical data. A graph is a sort of Graph through which statistical data are represented in the form of lines or curves drawn across the coordinated points plotted on its surface.

Graphs enable us in studying the cause and effect relationship between two variables. Graphs help to measure the extent of change in one variable when another variable changes by a certain amount.

Graphs also enable us in studying both time series and frequency distribution as they give clear account and precise picture of problem. Graphs are also easy to understand and eye catching.

#### **General Principles of Graphic Representation:**

There are some algebraic principles which apply to all types of graphic representation of data. In a graph there are two lines called coordinate axes. One is vertical known as Y axis and the other is horizontal called X axis. These two lines are perpendicular to each other. Where these two lines intersect each other is called '0' or the Origin. On the X axis the distances right to the origin have positive value (see fig. 7.1) and distances left to the origin have negative value. On the Y axis distances above the origin have a positive value and below the origin have a negative value.

#### **Methods to Represent a Frequency Distribution:**

Generally four methods are used to represent a frequency distribution graphically. These are Histogram, Smoothed frequency graph and Ogive or Cumulative frequency graph and pie diagram.

Kinds Graphical Representation:-

1. Pie Graph
2. Bar graph
3. Histogram
4. Frequency Polygon
5. Cumulative Frequency Curve
6. Cumulative Percentage Curve

### **1. Pie Graph**

A pie Graph is a type of graph that represents the data in the circular graph. A pie Graph requires a list of categorical variables and the numerical variables. Here, the term "pie" represents the whole and the "slices" represents the parts of the whole. In this article, we will discuss the definition of a pie Graph, its formula, an example to create a pie Graph, uses, advantages and disadvantages in detail.

What is Pie Graph?

The “**pie Graph**” also is known as “circle Graph” divides the circular statistical graphic into sectors or slices in order to illustrate the numerical problems. Each sector denotes a proportionate part of the whole. To find out the composition of something, Pie-Graph works the best at that time. In most of the cases, pie Graphs replace some other graphs like the bar graph, line plots, histograms etc.

### **Pie Graph Formula**

The pie Graph is an important type of data representation. It contains different segments and sectors in which each segment and sectors of a pie Graph forms a certain portion of the total(percentage). The total of all the data is equal to 360°.

**The total value of the pie is always 100%.**

To work out with the percentage for a pie Graph, follow the steps given below:

- Categorize the data
- Calculate the total
- Divide the categories
- Convert into percentages
- Finally, calculate the degrees

Therefore, the pie Graph formula is given as

$$\text{(Given Data/Total value of Data)} \times 360^\circ$$

Example :-

Imagine a teacher surveys her class on the basis of their favourite Sports:

Football	Hockey	Cricket	Basketball	Badminton
10	5	5	10	10

The data above can be represented by a pie-Graph as following and by using the circle graph formula i.e. the pie Graph formula given below. It makes the size of portion easy to understand.

**Step 1:** First, Enter the data into the table.

Football	Hockey	Cricket	Basketball	Badminton
10	5	5	10	10

**Step 2:** Add all the values in the table to get the total.

Total students are 40 in this case.

**Step 3:** Next, divide each value by the total and multiply by 100 to get a per cent:

<b>Football</b>	<b>Hockey</b>	<b>Cricket</b>	<b>Basketball</b>	<b>Badminton</b>
-----------------	---------------	----------------	-------------------	------------------

$(10/40) \times 100$ =25%	$(5/40) \times 100$ =12.5%	$(5/40) \times 100$ =12.5%	$(10/40) \times 100$ =25%	$(10/40) \times 100$ =25%
------------------------------	-------------------------------	-------------------------------	------------------------------	------------------------------

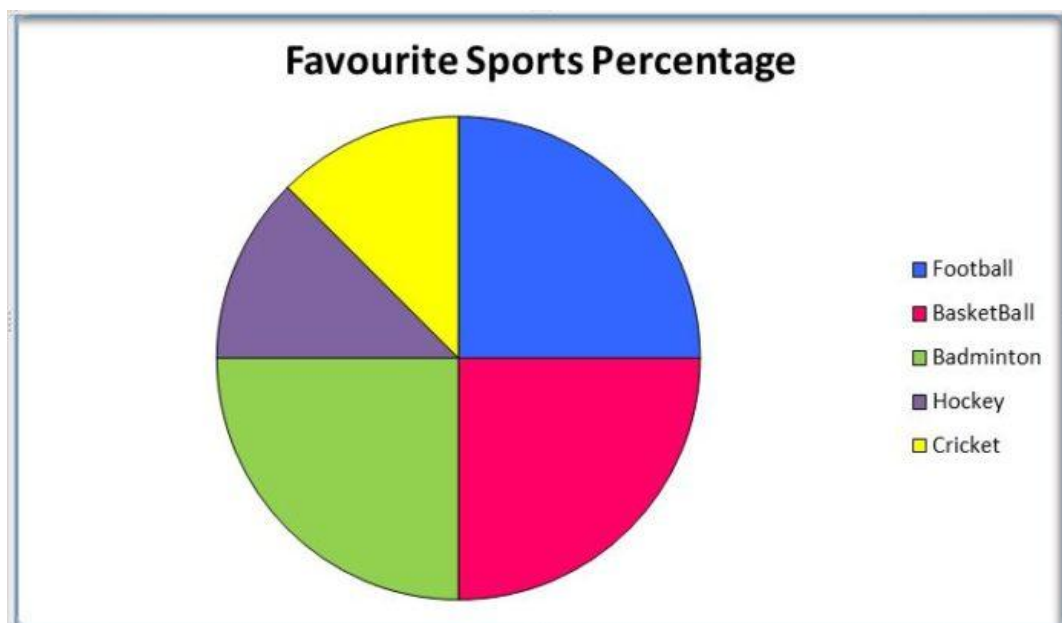
**Step 4:** Next to know how many degrees for each “pie sector” we need, we will take a full circle of 360° and follow the calculations below:

The central angle of each component = (Value of each component/sum of values of all the components)  $\times 360^\circ$

Football	Hockey	Cricket	Basketball	Badminton
$(10/40) \times 360^\circ$ =90°	$(5/40) \times 360^\circ$ =45°	$(5/40) \times 360^\circ$ =45°	$(10/40) \times 360^\circ$ =90°	$(10/40) \times 360^\circ$ =90°

Now you can draw a pie Graph.

**Step 5:** Draw a circle and use the protractor to measure the degree of each sector.



### Uses of Pie Graph

1. Pie diagram is useful when one wants to picture proportions of the total in a striking way.
2. When a population is stratified and each strata is to be presented as a percentage at that time pie diagram is used.

### Advantages of a Pie Graph

- The picture is simple and easy-to-understand
- Data can be represented visually as a fractional part of a whole
- It helps in providing an effective communication tool for the even uninformed audience

- Provides a data comparison for the audience at a glance to give an immediate analysis or to quickly understand information
- No need for readers to examine or measure underlying numbers themselves which can be removed by using this Graph
- To emphasize a few points you want to make, you can manipulate pieces of data in the pie Graph

### Disadvantages of a Pie Graph

- It becomes less effective, If there are too many pieces of data to use
- If there are too many pieces of data, and even if you add data labels and numbers may not help here, they themselves may become crowded and hard to read
- As this Graph only represents one data set, You need a series to compare multiple sets
- This may make it more difficult for readers when it comes to analyze and assimilate

### 2. Bar Diagram

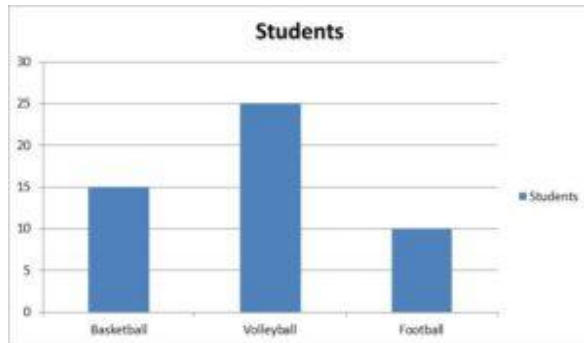
- Also known as a column graph, a bar graph or a bar diagram is a **pictorial representation of data**. It is shown in the form of rectangles spaced out with **equal spaces** between them and having **equal width**. The equal width and equal space criteria are important characteristics of a bar graph.
- Note that the height (or length) of each bar corresponds to the frequency of a particular observation. You can draw bar graphs both, vertically or horizontally depending on whether you take the frequency along the vertical or horizontal axes respectively. Let us take an example to understand how a bar graph is drawn.

Example:-Construct a Bar Graph from following data.

Students	No. of Students
Basketball	15
Volleyball	25
Football	10
Total = 50	

To above table depicts the number of students of a class engaged in any one of the three sports given. Note that the number of students is actually the frequency. So, if we take frequency to be represented on the y-axis and the sports on the x-axis, taking each unit on the y-axis to be equal to 5 students, we would get a graph that resembles the one below.

The blue rectangles here are called bars. Note that the bars have equal width and are equally spaced, as mentioned above. This is a simple bar diagram.



### 3.Histogram:

Histogram is a non-cumulative frequency graph, it is drawn on a natural scale in which the representative frequencies of the different class of values are represented through vertical rectangles drawn closed to each other. Measure of central tendency, mode can be easily determined with the help of this graph.

How to draw a Histogram:

1. Represent the class intervals of the variables along the X axis and their frequencies along the Y-axis on natural scale.
2. Start X axis with the lower limit of the lowest class interval. When the lower limit happens to be a distant score from the origin give a break in the X-axis to indicate that the vertical axis has been moved in for convenience.
3. Now draw rectangular bars in parallel to Y axis above each of the class intervals with class units as base: The areas of rectangles must be proportional to the frequencies of the corresponding classes.

Example :- Construct a Histogram Graph from following data.

C.I	f
20-24	2
25-29	2
30-34	5
35-39	10
40-44	6
45-49	2
50-54	3

**Solution:**

In this graph we shall take class intervals in the X axis and frequencies in the Y axis. Before plotting the graph we have to convert the class into their exact limits.

c.i.	f
19.5—24.5	2
24.5—29.5	2
29.5—34.5	5
34.5—39.5	10
39.5—44.5	6
44.5—49.5	2
49.5—54.5	3

Histogram plotted from the data.

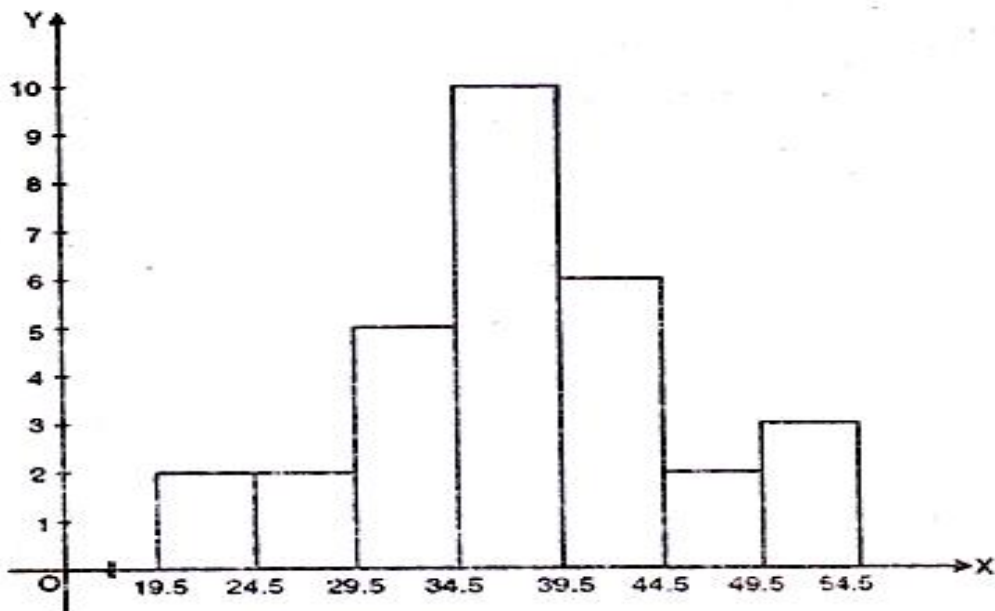


Fig. 7.2

**Advantages of histogram:**

1. It is easy to draw and simple to understand.
2. It helps us to understand the distribution easily and quickly.
3. It is more precise than the polygene.

**Limitations of histogram:**

1. It is not possible to plot more than one distribution on same axes as histogram.
2. Comparison of more than one frequency distribution on the same axes is not possible.

3. It is not possible to make it smooth.

**Uses of histogram:**

1. Represents the data in graphic form.
2. Provides the knowledge of how the scores in the group are distributed. Whether the scores are piled up at the lower or higher end of the distribution or are evenly and regularly distributed throughout the scale.
3. Frequency Polygon. The frequency polygon is a frequency graph which is drawn by joining the coordinating points of the mid-values of the class intervals and their corresponding frequencies.

**4.Frequency Polygon**

1. Draw a horizontal line at the bottom of graph paper named 'X' axis. Mark off the exact limits of the class intervals along this axis. It is better to start with c.i. of lowest value. When the lowest score in the distribution is a large number we cannot show it graphically if we start with the origin. Therefore put a break in the X axis ( ) to indicate that the vertical axis has been moved in for convenience. Two additional points may be added to the two extreme ends.
2. Draw a vertical line through the extreme end of the horizontal axis known as OY axis. Along this line mark off the units to represent the frequencies of the class intervals. The scale should be chosen in such a way that it will make the largest frequency (height) of the polygon approximately 75 percent of the width of the figure.
3. Plot the points at a height proportional to the frequencies directly above the point on the horizontal axis representing the mid-point of each class interval.
4. After plotting all the points on the graph join these points by a series of short straight lines to form the frequency polygon. In order to complete the figure two additional intervals at the high end and low end of the distribution should be included. The frequency of these two intervals will be zero.

**Draw a frequency polygon from the following data:**

Marks in Mathematics	40-45	45-49	50-54	55-59	60-64	65-69	70-74	75-79	80-84	85-89	90-95	95-99
No. of students	1	3	2	4	5	6	10	8	5	6	2	1

**Solution:**

In this graph we shall take the class intervals (marks in mathematics) in X axis, and frequencies (Number of students) in the Y axis. Before plotting the graph we have to convert the c.i. into their exact limits and extend one c.i. in each end with a frequency of 0.

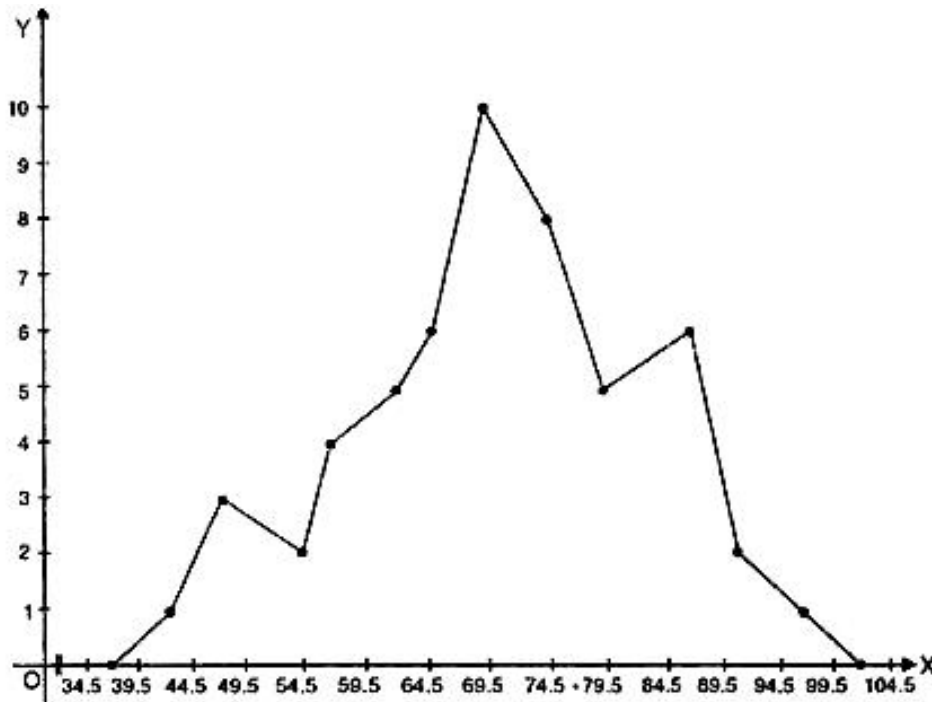
**Class intervals with exact limits:**

Fig. 7.3 Frequency polygon plotted from the data.

**Advantages of frequency polygon:**

1. It is easy to draw and simple to understand.
2. It is possible to plot two distributions at a time on same axes.
3. Comparison of two distributions can be made through frequency polygon.
4. It is possible to make it smooth.

**Limitations of frequency polygon:**

1. It is less precise.
2. It is not accurate in terms of area the frequency upon each interval.

**Uses of frequency polygon:**

1. When two or more distributions are to be compared the frequency polygon is used.
2. It represents the data in graphic form.
3. It provides knowledge of how the scores in one or more group are distributed. Whether the scores are piled up at the lower or higher end of the distribution or are evenly and regularly distributed throughout the scale.



## 5.Cumulative Frequency Curve

**Meaning:-**This is a diagram that displays cumulative frequency.

Marks in History	0—9	10—19	20—29	30—39	40—49	50—59	60—69	70—79	80—89	90—99
No. of Students	3	5	9	12	18	17	10	3	2	1

### Solution:

To plot this graph first we have to convert, the class intervals into their exact limits. Then we have to calculate the cumulative frequencies of the distribution.

. c.i.	f	c.f. (cumulative frequencies)
0—9.5	3	3
9.5—19.5	5	8
19.5—29.5	9	17
29.5—39.5	12	29
39.5—49.5	18	47
49.5—59.5	17	64
59.5—69.5	10	74
69.5—79.5	3	77
79.5—89.5	2	79
89.5—99.5	1	80

Now we have to plot the cumulative frequencies in respect to their corresponding class-intervals.

**Ogive plotted from the data given above:**

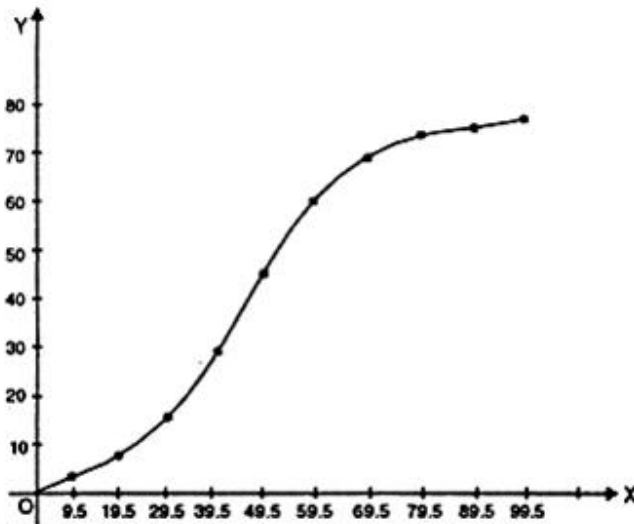


Fig. 7.5.

### Uses:

1. Ogive is useful to determine the number of students below and above a particular score.
2. When the median as a measure of central tendency is wanted.

3. When the quartiles, deciles and percentiles are wanted.
4. By plotting the scores of two groups on a same scale we can compare both the groups.

## 6. Cumulative percentage Curve

Cumulative percentage is another way of expressing frequency distribution. It calculates the percentage of the cumulative frequency within each interval, much as relative frequency distribution calculates the percentage of frequency.

The main advantage of cumulative percentage over cumulative frequency as a measure of frequency distribution is that it provides an easier way to compare different sets of data.

Cumulative frequency and cumulative percentage graphs are exactly the same, with the exception of the vertical axis scale. In fact, it is possible to have the two vertical axes, (one for cumulative frequency and another for cumulative percentage), on the same graph.

Cumulative percentage is calculated by dividing the cumulative frequency by the total number of observations ( $n$ ), then multiplying it by 100 (the last value will always be equal to 100%). Thus,

$$\text{cumulative percentage} = (\text{cumulative frequency} \div n) \times 100$$

### Example 1 – Calculating cumulative percentage

#### Example:-

For 25 days, the snow depth at Whistler Mountain, B.C. was measured (to the nearest centimeter) and recorded as follows:

242, 228, 217, 209, 253, 239, 266, 242, 251, 240, 223, 219, 246, 260, 258, 225, 234, 230, 249, 245, 254, 243, 235, 231, 257.

#### Answers:

1. The snow depth measurements range from 209 cm to 266 cm. In order to produce the table, the data are best grouped in class intervals of 10 cm each. In the *Snow depth* column, each 10-cm class interval from 200 cm to 270 cm is listed. The *Frequency* column records the number of observations that fall within a particular interval. This column represents the observations in the *Tally* column, only in numerical form. Each of the numbers in the *Endpoint* column is the highest number in each class interval. In the interval of 200 cm to 210 cm, the endpoint would be 210. The *Cumulative frequency* column lists the total of each frequency added to its predecessor, as seen in the exercises in the previous section.

The *Cumulative percentage* column divides the cumulative frequency by the total number of observations (in this case, 25). The result is then multiplied by 100. This calculation gives the cumulative percentage for each interval.

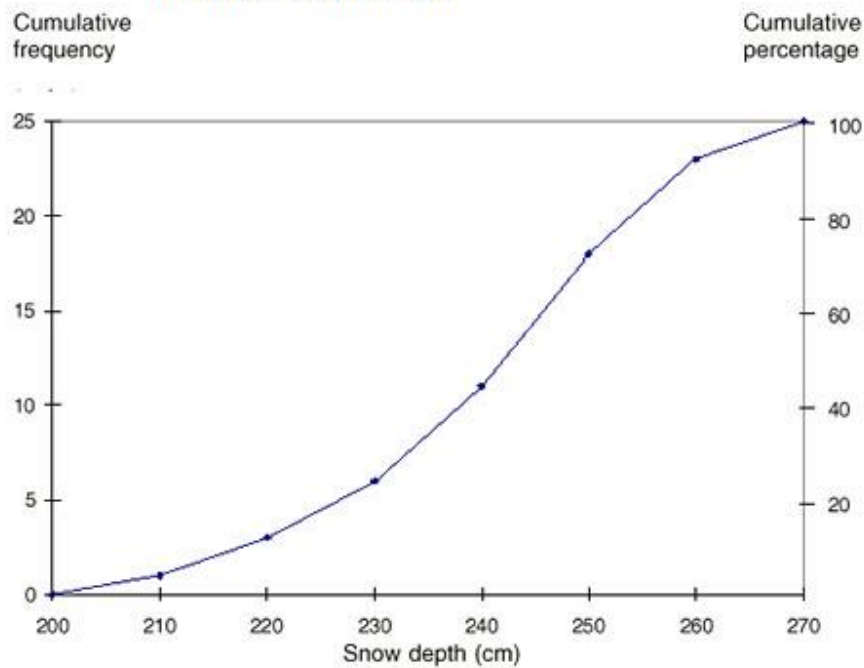
Snow depth (x)	Frequency (f)	Endpoint	Cumulative frequency	Cumulative percentage
		200	0	$0 \div 25 \times 100 = 0$
200 to 210	1	210	1	$1 \div 25 \times 100 = 4$
210 to 220	2	220	3	$3 \div 25 \times 100 = 12$
220 to 230	3	230	6	$6 \div 25 \times 100 = 24$
230 to 240	5	240	11	$11 \div 25 \times 100 = 44$
240 to 250	7	250	18	$18 \div 25 \times 100 = 72$
250 to 260	5	260	23	$23 \div 25 \times 100 = 92$
260 to 270	2	270	25	$25 \div 25 \times 100 = 100$

2. Apart from the extra axis representing the cumulative percentage, the graph should look exactly the same as that drawn in [Example 2](#) of the section on Cumulative frequency. The *Cumulative percentage* axis is divided into five intervals of 20, while the *Cumulative frequency* axis is divided into five intervals of 5. The *Snow depth* axis is divided by the endpoints of each 10-cm class interval.

Using each endpoint to plot the graph, you will discover that both the cumulative frequency and the cumulative percentage land in the same spot. For example, using the endpoint of 260, plot your point on the 23rd day (cumulative frequency). This point happens to be in the same place where the cumulative percentage (92%) will be plotted. You have to be very careful when you are building a graph with two y-axes. For example, if you have 47 observations, you might be tempted to use intervals of 5 and end your y-axis at the cumulative frequency of 50. However, when you draw your y-axis for the cumulative percentage, you must put the 100% interval at the same level as the 47 mark on the other y-axis—not at the 50 mark. For this example, a cumulative frequency of 47 represents 100% of your data. If you put the 100% at the top of the scale where the 50 interval is marked, your line for the cumulative frequency will not match the line for the cumulative percentage.

The plotted points join to form an ogive, which often looks similar to a stretched **S**. Ogives are used to determine the number, or percentage, of observations that lie above or below a specified value. For example, according to the table and the graph, 92% of the time the snow depth recorded in the 25-day period was below the 260 cm mark.

**Figure 1. Snow depth measured at Whistler Mountain, B.C., 25-day period**



The following information can be gained from either the graph or table:

- during the 25-day period, 24% of the time the recorded snow depth was less than 230 cm
- on 7 of the 25 days, snow depth was at least 250 cm