1. Introduction to Frequency Polygon

Welcome to the series of e-Learning modules on frequency polygon and ogive. In this module we are going to cover the definition, utility, construction and analysis of frequency polygon and ogive as a graphical tool of representation of data.

By the end of this session, you will be able to:

- Explain frequency polygon and ogive
- Explain the uses of frequency polygon and ogive
- Explain how to do the construction of frequency polygon and ogive from a given data
- Explain how to do the analysis and interpretation of the data

What is a frequency polygon?
Frequency Polygon has been derived from the word "polygon" which means many sides.

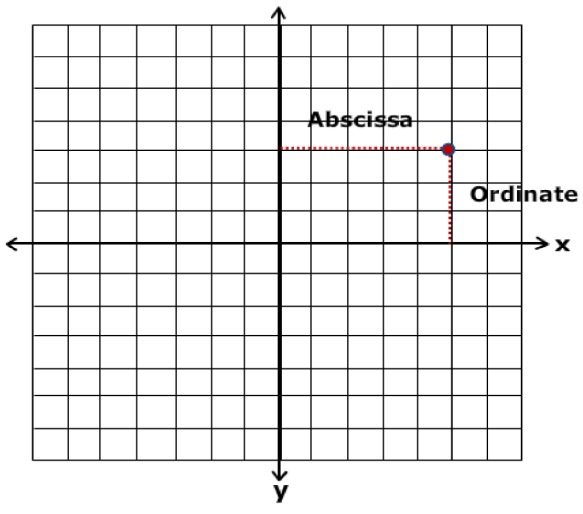
In statistics, it means a graph of a frequency distribution. A frequency polygon is obtained from a histogram by joining the mid-points of the top of various rectangles with the help of straight lines.

The total area under the polygon remains equal to the area under histogram and two arbitrary Classes, each with zero frequency are added on both ends.

Frequency means the number of ‘occurrences’ of the data and polygon means ‘many angled diagram.’ A frequency polygon is a graphical representation of the frequency distribution. It is formed by connecting a series of points.

The abscissa of each point is the midpoint of the interval in which the point lies. The ordinate of each point is the frequency for the interval.
The polygon is closed at each end by drawing a line from the endpoints to the horizontal axis at the midpoint of the next interval.

Figure 1



Definition:
“Frequency polygons are a graphical device for understanding the shapes of distributions.”

There are two types of frequency polygon.

- General Frequency Polygon
- Relative Frequency Polygon

General frequency polygons are polygons where the actual numbers of the data points are retained for comparison.
Relative frequency polygons are polygons where the numbers of the data points are converted into percentage frequencies for comparison.

Let us see, when we should use a frequency polygon.

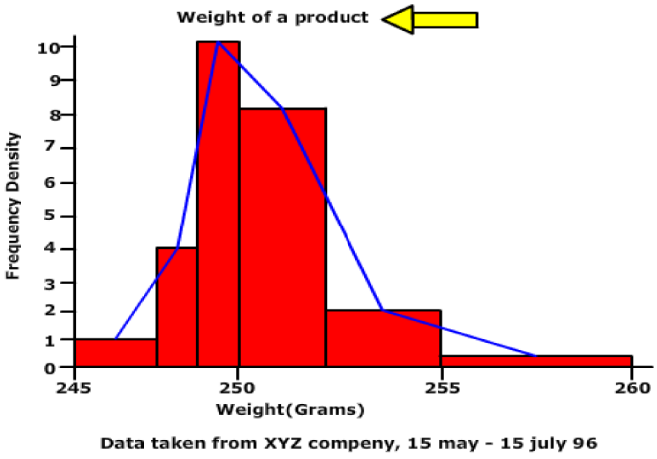
- Frequency polygons are useful for effective comparing of two or more distributions by overlaying the frequency polygons drawn for different data sets
- Frequency polygons are used for easy ascertainment of the value of mode of the data by drawing a perpendicular from the apex of the polygon to the X-axis

2. Uses and Construction of Frequency Polygon

Parts of frequency polygon are:

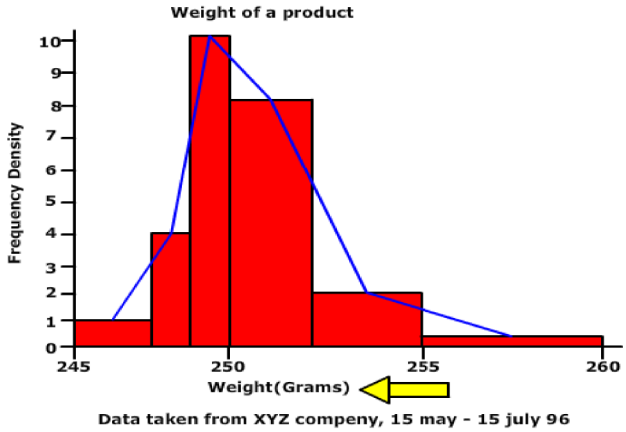
1. Title: The title briefly describes the information that is contained in the frequency polygon.

Figure 2



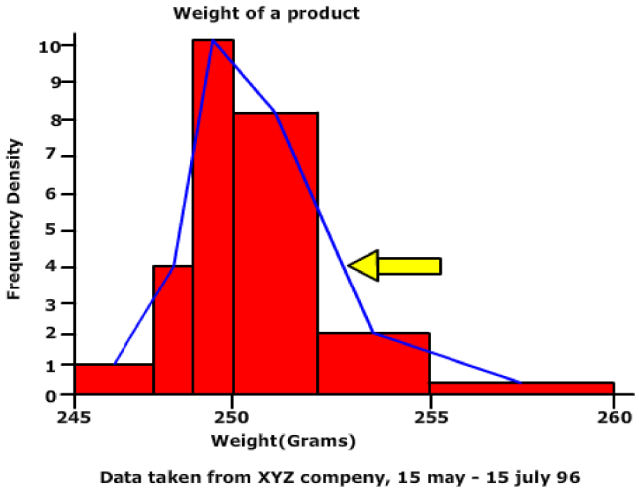
2. Horizontal or X-Axis: The horizontal or X-axis shows you the scale of values into which the measurements fit. These measurements are generally grouped into intervals to help you summarize large data sets and the mid values of the intervals are considered.

Figure 3



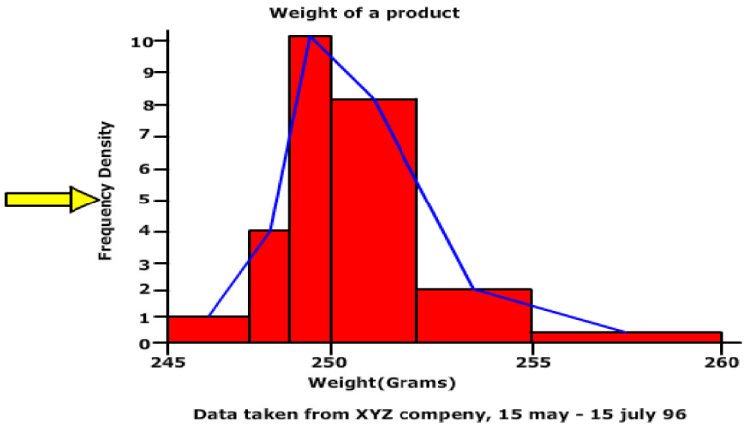
3. Line: The line in a polygon is drawn in such a way that it covers the total area under the histogram.

Figure 4



4. Vertical or Y-Axis: The vertical or Y-axis is the scale that shows you the number of times the values within an interval occurred. The number of times the value occurred is referred to as "frequency."

Figure 5



5. Legend: The legend provides additional information like where the data came from and how the measurements were gathered.

Figure 6

Let us see how the construction of frequency polygon is done from a given data.
A frequency polygon is constructed in two ways one with histogram and the other without the histogram. In either case the following steps are followed in general.

Step 1:
Organizing the data into groups: In this step the collected data is arranged in ascending order and the largest and the smallest value of the data is identified to find the range of the data which helps in creating the groups called the class intervals of the data.

Step 2:
Preparation of a frequency table: In this step the data organized is represented in the form of a table with two columns one representing the class interval and the other representing the frequency that is indicating the number of data points in each class interval.

Figure 7

Class Interval	Frequency

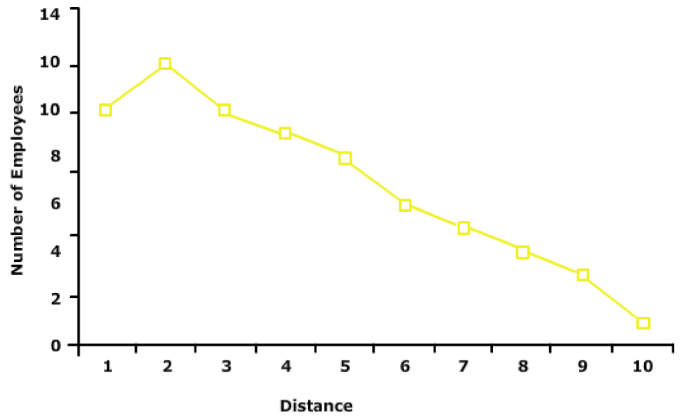
Step 3:
Calculating the mid points of the group: In this step we add a third column to the table called mid-point which helps in identifying the mid value of each class. The mid value is calculated by adding up the boundary limits of each class interval and dividing it by two.

Figure 8

Class Interval	Frequency	Mid-point

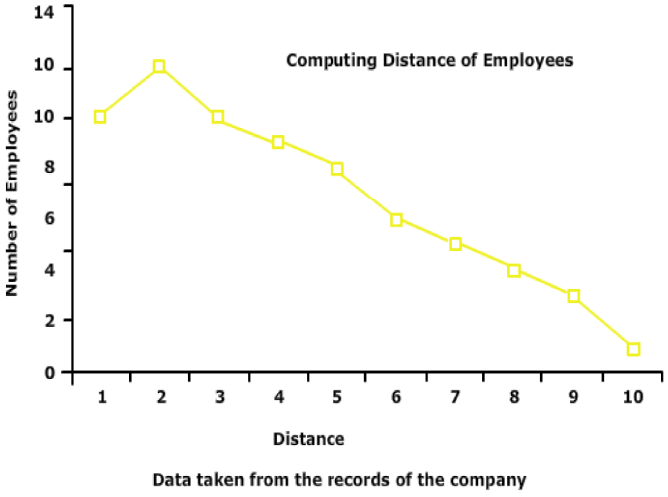
Step 4:
Plotting the graph: In this step the data is plotted in a graph where the abscissa of each point is the midpoint of the interval in which the point lies. The ordinate of each point is the frequency for the interval. The polygon is closed at each end by drawing a line from the endpoints to the horizontal axis at the midpoint of the next interval.

Figure 9



Step 5:
Add the title & legend: In this step we try to give a title and legend to the graph so that it helps in identifying the five W of the data (Who, Where, What, When and Why) that are important for understanding and interpreting the data. The legend can have information like the sample size or the time period or the source of the data collected for easy clarification or description of the information of the data.

Figure 10



Whenever the frequency polygon is constructed with a histogram the mid values are not calculated.

1. We prepare the histogram for the given data by taking the class intervals on the X-axis (abscissa) and the frequencies on the Y-axis the coordinate.
2. The rectangle bars are drawn and joined by a straight line connecting the mid-points of the upper horizontal side of each rectangle.
3. Both the ends are extended to the base line by creating two hypothetical classes at each end with zero frequency.

3. Illustration of Frequency Polygon

Illustration for frequency polygon.
Following are the marks obtained by 20 students of Class A in statistics paper for 50 marks.

Figure 11

10	21	22	25	27	29	32	35	36	32
35	36	39	39	42	42	46	48	49	50

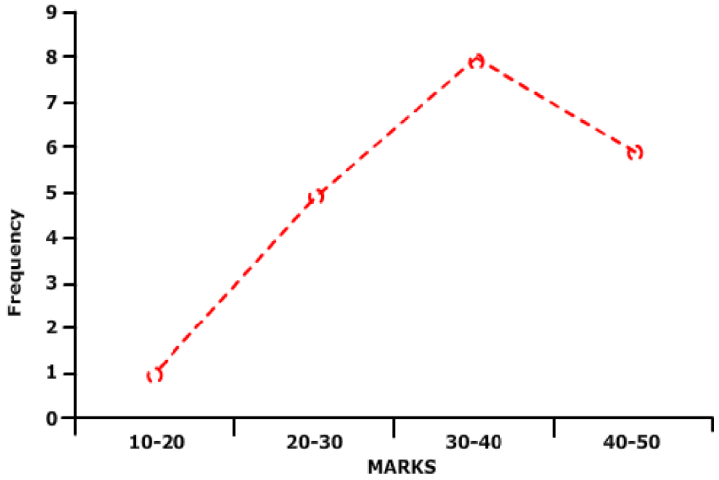
Step 1: Organizing the data into groups:
In this step the largest value of the data is 50 and the least value is 10 so the range of the data will be 50-10=40 divided by 4 class intervals we will get a width of 10 so the class intervals will be 10-20, 20-30, 30-40, 40-50.

Step 2: Preparation of a frequency table:
Here, for the class interval 10-20 frequency is 1.
For the class interval 20-30 frequency is 5.
For the class interval 30-40 frequency is 8.
For the class interval 40-50 frequency is 6.

Step 3: Calculating the mid points of the group:
In this step we calculate the mid points by adding the class boundaries and dividing it by 2. That is in the first case 10+20=30. The total is divided by 2 i.e. 30/2 = 15. In a similar way it is calculated for all the class intervals and the new table with three columns is formed. Here the midpoints are 15, 25, 35 and 45.

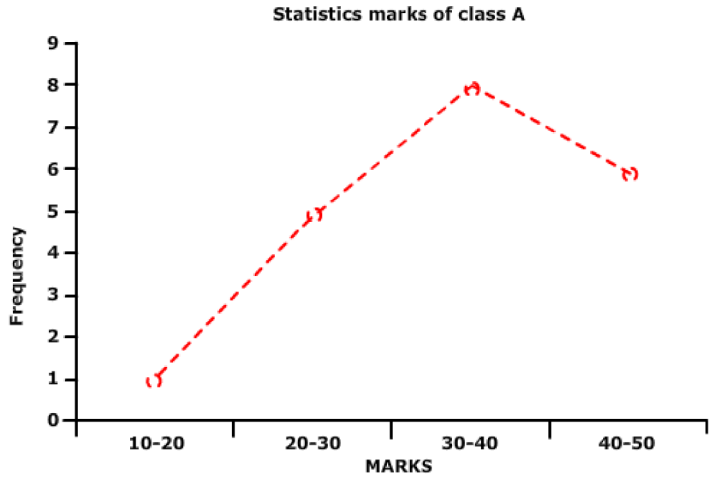
Step 4: Plotting the graph:
In this step we take the mid points in the X-axis and the frequency in the Y-axis and then plot the frequencies against the mid points. The points so obtained are joined by straight lines and a frequency polygon is formed.

Figure 12



Step 5: Add title & legend: Here we give a title to the graph as Statistics marks of class A the legend in this case could be the marks obtained by the students in the internal test for March 2012. This explains what and when of the data.

Figure 13



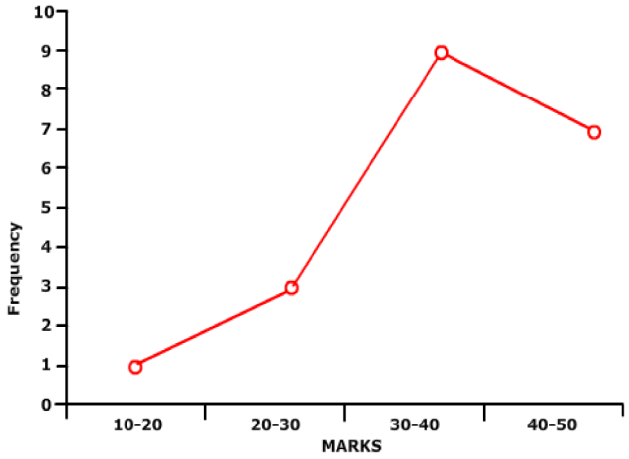
The frequency polygon is used for comparison of data. So for a detail understanding let us add another ready data for the same illustration that is marks of 20 students of class B in statistics. Then we get a new table as follows:

Figure 14

Marks (Class B)	Frequency	Mid-point
10-20	1	15
20-30	3	25
30-40	9	35
40-50	7	45

The above table is also plotted in the same graph now we can compare the performance of the students of two classes A & B in statistics and arrive at conclusions.
Hence, while interpreting the data from a frequency polygon we need to keep in mind the purpose or the information we need to arrive at from the data.

Figure 15



Well, in this case, these are the marks of students in two classes got on an exam. What might someone be interested from this information? Well, the fact that there are two classes should suggest something to you – you could compare how the two classes went on the exam.
So, the bigger the mark, the better a student has done right? A high mark out of 50 is good, and a low mark out of 50 is bad! Okay, so let's look at the frequency polygon.

Notice how Class A had more students with low marks – the dashed line (representing Class A) is higher than the solid line (representing Class B) in the left hand area of the graph. This tells us that Class A has more students who performed poorly than Class B.
What about on the right hand side of the graph, in the higher marks section? Well, in this section the solid line (representing Class B) is higher than the dashed line (representing Class A). This tells us that Class B had more students with high marks on the exam.
So this diagram tells us what? Well, Class B has fewer students with low marks and more students with high marks when compared with Class A. This means that Class B went better on the exam than Class A.

So you might say something like from looking at the frequency polygons:

- Class A has more students who scored a low mark on the exam than Class B
- Class B has more students who scored a high mark on the exam than Class A
- This tells us that Class B performed better on the exam than Class A

Advantages and Disadvantages of Frequency polygon.
Advantages:

- It is simple for construction.
- Sketches an outline of the data pattern.
- When several distributions are plotted there is a smooth curve like

structure.

- Several distributions can be plotted on the same axis.
- Comparison of data is easily done.

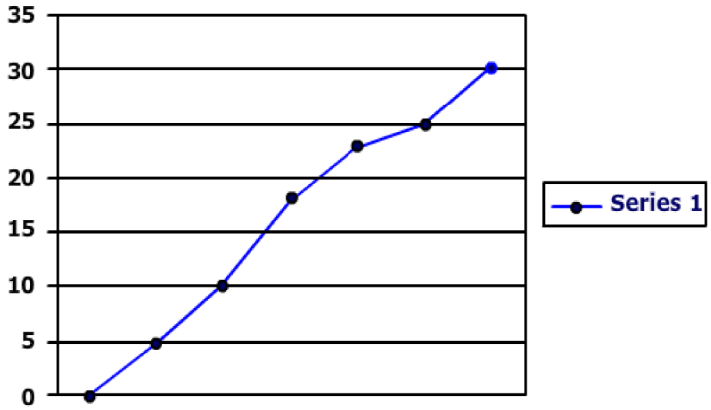
Disadvantages:

- It is difficult to construct frequency polygons for open ended classes.
- It can be misleading if the distribution has unequal class intervals and suitable adjustments in frequencies are not made.

4. Introduction and Construction of Ogives

Data may be expressed using a single line. An ogive (a cumulative line graph) is best used when you want to display the total at any given time. The relative slopes from point to point will indicate greater or lesser increases; for example, a steeper slope means a greater
Let us see when we should use an ogive.
If you simply want to keep track of a total and your individual values are periodically combined, an ogive is an appropriate display.

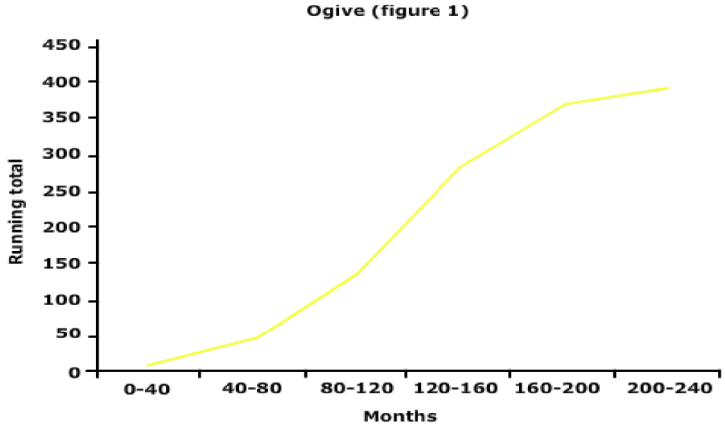
Figure 16



For example, if you saved Rs.300 in both January and April and Rs.100 in each of February, March, May, and June, an ogive would look like as in the figure below. An ogive displays a running total.

Although each individual month's savings could be expressed in a bar chart as shown in Figure ,

Figure 17

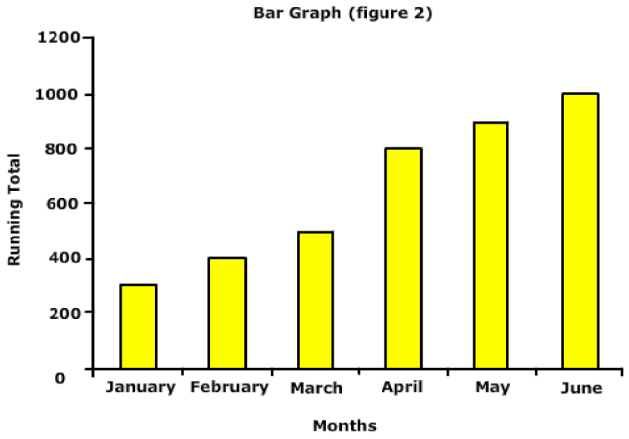


We can easily see the amount of total growth or loss in the ogive curve and not in a bar graph.

Utility of ogives:

- To determine or portray the number of values above or below a given point
- To compare two or more distributions
- To determine certain values graphically like the median, quartiles, percentiles, deciles, etc.

Figure 18



Role of cumulative frequency in ogive:
Cumulative frequency is defined as a running total of frequencies. The frequency of an element in a set refers to-how many elements are there in the set. Cumulative frequency can also defined as the sum of all previous frequencies up to the current point.

The cumulative frequency is important when analysing data, where the value of the cumulative frequency indicates the number of elements in the data set that lie below the current value. The cumulative frequency is also useful when representing data using diagrams like histograms.

Now let us see how to construct Ogives from the given data.
There are two methods of cumulating a series, and based on both the methods we construct two ogives one is a less than ogive and the other is a more than ogive.

First method
The frequencies of all preceding classes are added to the frequency of a class. This series is called the less than cumulative series.
It is constructed by adding the first class frequency to the second class frequency and then to the third class frequency and so on. The downward cumulating result in the less than cumulative series.

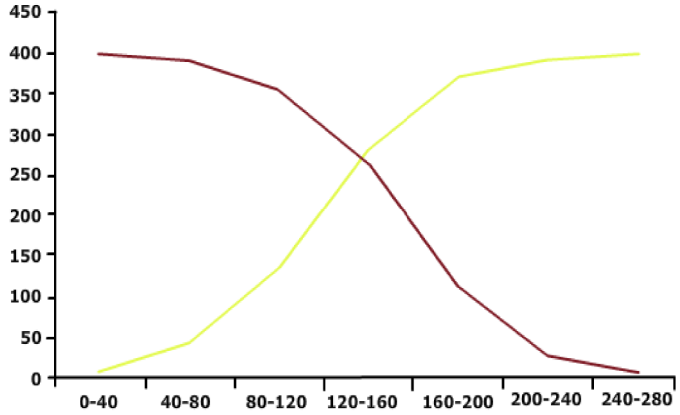
Second method
The frequencies of the succeeding classes are added to the frequency of a class. This series is called as the more than or greater than cumulative series.
It is constructed by subtracting the first class, second class frequency from the total, third

class frequency from that and so on. The upward cumulating result is greater than or more than the cumulative series.

When the graphs of these series are drawn, we get a cumulative frequency curve or Ogives.

Points to remember:
1. The less than cumulative frequency curve is known as Less than Ogive and the greater than cumulative frequency curve is known as the Greater than or More than Ogive.
2. Less than Ogive curves are obtained by plotting less than cumulative frequencies against the upper limits of each class interval.

Figure 19



3. More than Ogive curves are obtained by plotting more than cumulative frequencies against the lower limits of each class interval.
4. Less than cumulative frequency curve slope upwards from left to right where as more than cumulative curve slope downwards from left to right.
5. Ogives are the graphical representations used to find the Median of a frequency distribution.

Steps for constructing a less than Ogive chart (less than Cumulative frequency curve):
1. Draw and label the horizontal and vertical axes.
2. Take the cumulative frequencies along the y axis (vertical axis) and the upper class limits on the x axis (horizontal axis)
3. Plot the cumulative frequencies against each upper class limit.
4. Join the points with a smooth curve.

Steps for constructing a greater than or more than Ogive chart (more than Cumulative frequency curve):
1. Draw and label the horizontal and vertical axes.
2. Take the cumulative frequencies along the y axis (vertical axis) and the lower class limits on the x axis (horizontal axis)
3. Plot the cumulative frequencies against each lower class limit.
4. Join the points with a smooth curve.

How to interpret an ogive.
The point of intersection of less than cumulative frequency curve and greater than or more than cumulative frequency curve is the Median of the distribution. So we can find the

middlemost value of the series by drawing a perpendicular to the X-axis.

5. Advantages, Disadvantages and Illustration of Ogives

Advantages & Disadvantages of Ogive.
Advantages of Ogives are:

- Good visual representation of raw data
- Helps in checking accuracy and fairness of calculations
- Clearly defines each interval in the frequency distribution
- Helps in clarification of changes between two sets of data
- Has a wide usage in business and media as it helps in identifying the proportion of data points above or below a particular value

Disadvantages of Ogives are:

- Complicated in construction
- It fails to reflect all the data points of the data set
- Does not highlight about other measures of central tendency or dispersion or skewness
- Needs usage of additional words written or verbal for explanation
- It cannot describe the qualitative data like behavior, condition of interest, etc.
- Does not explain the key assumptions, norms cause and effects of the data
- Gives a false impression in the minds of the viewer

Example:
Let us take an example to understand the calculation of the cumulative frequencies and construction of both the less than and more than ogives.

From the following data draw a less than and more than ogive curve.

Figure 20

Salary per day (Rs.)	0-40	40-80	80-120	120-160	160-200	200-240	240-280
Frequency	9	36	91	147	87	22	8

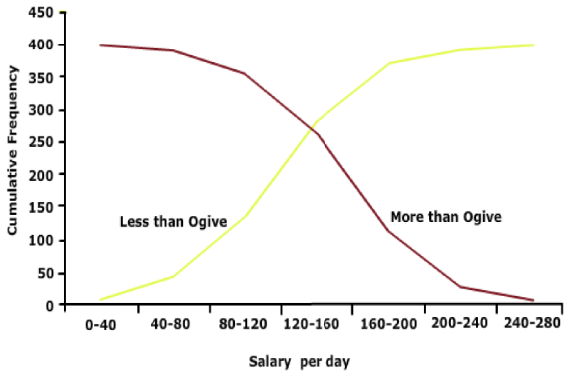
First let us prepare the cumulative frequency table for both less than and more than ogives.
Less than ogive:
While calculating the cumulative frequency in case of less than, the frequency of the first class remains the same then we go on adding the frequency of the corresponding class like

9+36=45, 45+91=136, and so on. In this case we plot the cumulative frequency against the upper limit of the class interval.

More than ogive:
While calculating the cumulative frequency in case of more than, we start from the last class interval, the frequency of the last class interval remains the same and thereon we add the frequencies from bottom to top like 8+22=30, 30+87=117, and so on we get the following table. In this case we plot the cumulative frequency against the lower limit of the class interval.

We get the following diagram with two curves intersecting each other. The perpendicular drawn from the point of intersection to the X-axis gives us the value of the median of the data. Also we can analyze and find out how many workers are drawing less than Rs.180 per day and more than Rs.180 per day.

Figure 21



To find a solution we can draw two perpendiculars from the Y axis to touch the more curve and the less than curve corresponding to the salary of Rs.180 and we find that the number of workers drawing salary less than Rs.180 is 73 and more than Rs. 180 is 327.

Here's a summary of our learning in this session:

- Definition of frequency polygon and ogives
- Usage of frequency polygon and ogive
- Steps involved in calculation and construction of frequency polygon and ogive
- Advantages and limitations of frequency polygon and ogive