

Chapter 1:

Fundamentals

- Variates
- Frequency distribution
- Simple frequency distribution
- Grouped frequency distribution
- some definitions
- Cumulative frequency distribution

Variates

- A characteristic which can be expressed numerically is called a *variate* or *variable* (symbol is x , y , u or X).
- The number of times each variate occurs is known as its *frequency* (symbol is f or F).
- A frequency table is a chart consisting of the variates with their respective frequencies.
- A classification showing different values of a variate and the corresponding frequency is known as *frequency distribution*.

Frequency
Distribution

```
graph TD; A[Frequency Distribution] --> B[Simple Frequency Distribution]; A --> C[Grouped Frequency Distribution];
```

Simple
Frequency
Distribution

Grouped
Frequency
Distribution

Simple Frequency distributions

This shows the values of the variates are individual.

For example, the table below:

Table I.

<i>Weekly wages (Rs.)</i>	<i>No. of workers</i>	<i>Weekly wages (Rs.)</i>	<i>No. of workers</i>
9	2	30	4
10	3	32	5
17	4	35	4
18	2	37	3
19	1	40	4
20	4	41	2
25	3	45	7
26	3	47	1
28	5	50	3

Grouped Frequency Distribution

For huge number of figures, say, 1000, the above method of arrangement will not be helpful.

In such cases, the values of the variate may be shown in groups or intervals giving rise to a grouped frequency distribution, as shown below:

Table II. Grouped Frequency Distribution

<i>Variate (Weekly Wages) (in Rs.)</i>	<i>Frequency (No. of workers)</i>
From 1 to 10	5
„ 11, 20	11
„ 21, 30	15
„ 31, 40	16
„ 41, 50	13
Total	60

Few Terms

(Associated with grouped frequency distribution)

- (a) Class interval
- (b) Class frequency,
total frequency
- (c) Class-limits
(upper and lower).
- (d) Class-boundaries
(upper and lower)
- (e) Mid value of class-interval.
- (f) Width of class-interval
- (g) Frequency density
- (h) Percentage frequency

(a) Class Interval.

A large number of observations having a wide range, is usually classified in several groups.

Each of these groups is known as class-interval (or class).

(b) Class frequency, Total frequency.

The number of observations (frequency) in a particular class interval is known as *class frequency*.

The sum of all frequencies is called the *total frequency*.

<i>Class- Intervals</i>	<i>Class- frequency</i>
(1)	(2)
1 – 10	5
11 – 20	11
21 – 30	15
31 – 40	16
41 – 50	13
Total	60

(c) Class limits.

The two ends of a class interval are called class limits.

Of a particular class interval, the smaller and greater numbers are known as lower and upper-class limits respectively.

<i>Class- Intervals</i>	<i>Class- frequency</i>	<i>Class-limits</i>	
		<i>lower</i>	<i>upper</i>
(1)	(2)	(3)	(4)
1 – 10	5	1	10
11 – 20	11	11	20
21 – 30	15	21	30
31 – 40	16	31	40
41 – 50	13	41	50
Total	60	—	—

(d) Class-boundaries.

$$\text{lower class boundary} = \text{lower class limit} - \frac{1}{2}d,$$

$$\text{upper class boundary} = \text{upper class limit} + \frac{1}{2}d,$$

where

d = common difference between upper class of any class interval with the lower class of the next class-interval.

<i>Class- Inervals</i>	<i>Class- frequency</i>	<i>Class-boundaries</i>	
(1)	(2)	<i>lower</i>	<i>upper</i>
(1)	(2)	(5)	(6)
1 – 10	5	0.5	10.5
11 – 20	11	10.5	20.5
21 – 30	15	20.5	30.5
31 – 40	16	30.5	40.5
41 – 50	13	40.5	50.5
Total	60	—	—

(e) Mid value (or class mark).

The value exactly at the middle of a class-interval is known as its mid value.

It is calculated by adding the two class limits and divided by 2 (or, adding the two class-boundaries divided by 2).

<i>Class- Intervals</i>	<i>Class- frequency</i>	<i>Mid-value</i>
(1)	(2)	(7)
1 – 10	5	5.5
11 – 20	11	15.5
21 – 30	15	25.5
31 – 40	16	35.5
41 – 50	13	45.5
Total	60	—

(f) Width of class interval.

The width (or size) of a class-interval is the difference between the class boundaries (not class limits).

$$\text{width} = \text{upper class boundary} - \text{lower class boundary}$$

<i>Class- Intervals</i>	<i>Class- frequency</i>	<i>Width of class</i>
(1)	(2)	(8)
1 – 10	5	10
11 – 20	11	10
21 – 30	15	10
31 – 40	16	10
41 – 50	13	10
Total	60	—

(g) Frequency density.

It is the ratio of the class frequency to the width of that class interval

$$\textit{Frequency density} = \frac{\textit{Class frequency}}{\textit{Width of the class}}$$

<i>Class- Intervals</i>	<i>Class- frequency</i>	<i>Frequency density</i>
(1)	(2)	(9)
1 – 10	5	0.5
11 – 20	11	1.1
21 – 30	15	1.5
31 – 40	16	1.6
41 – 50	13	1.3
Total	60	—

(h) Percentage Frequency.

It is the ratio of class-frequency to total frequency expressed as percentage

$$\textit{Percentage Frequency} = \frac{\textit{Class frequency}}{\textit{Total frequency}} \times 100$$

<i>Class- Intervals</i>	<i>Class- frequency</i>	<i>Percentage frequency</i>
(1)	(2)	(10)
1 – 10	5	8.33
11 – 20	11	18.33
21 – 30	15	25.00
31 – 40	16	26.67
41 – 50	13	21.67
Total	60	100.00

Cumulative frequency distribution

It is a form of frequency distribution in which each frequency beginning with the second (from the top) is added with the total of the previous ones, the class-intervals being adjusted accordingly.

<i>Marks</i>	<i>Frequency</i>	<i>Cumulative frequency</i>
25–29	7	7
30–34	10	17 (= 10 + 7)
35–39	13	30 (= 13 + 17)
40–44	8	38 (= 8 + 30)
45–49	13	51 (= 13 + 38)
50–54	6	57 (= 6 + 51)
55–59	3	60 (= 3 + 57)
60–64	3	63 (= 3 + 60)
65–69	1	64 (= 1 + 63)
70–74	1	65 (= 1 + 64)
Total	65	—

This kind of cumulative frequency is known as 'less than type' cumulative frequency when addition is done from top.

Conversely if addition is done from below, then it will be 'greater than type' cumulative frequency.

Example. From the following table find (a) the less than and (b) greater than cumulative frequencies,

(c) cumulative frequency distribution, (d) cumulative percentage distribution.

Wages (Rs.)	11-20	21-30	31-40	41-50	51-60	60-70	Total
Frequency	5	7	12	15	8	3	50

<i>Wages (Rs.)</i>	<i>Cumulative Frequency</i>		<i>Cumulative percentage (less than)</i>
	<i>(less than)</i>	<i>(greater than)</i>	
10.5	0	50	0
20.5	5	45	10
30.5	12	38	24
40.5	24	26	48
50.5	39	11	78
60.5	47	3	94
70.5	50	0	100

Graphical Representation of Frequency Distribution

It is the representation or presentation of data as Diagrams and Graphs.

Advantages of Graphical Representation of Data

1. Data are presented pictorially.
2. Give better insight and understanding of the data.
3. Makes the presentation eye-catching.
4. The data become more logical (clear).
5. The comparison becomes easy.
6. Can derive the conclusion from data very quickly.
7. Give the spread of the data.
8. Reduce space for data representation.

Disadvantages of Graphical Representation of Data

1. A graph cannot represent all details of the variables.
2. Very difficult to include and study the small differences in large measurements.
3. Graphs usually show approximate figures.
4. Graphs are only a supplement to the tabular presentation of data.
5. Graphs cannot be an alternative to tabular presentation.

Methods of Graphical Representation of Data

- (1). Line Diagram
- (2). Bar Diagram
- (3). Histogram
- (4). Frequency Polygon
- (5). Frequency Curve
- (6). Pie Chart (Circle Diagram)
- (7). Ogive

(1). Line Diagram

1. The line diagram is the simplest method of graphical representation.
2. In line diagram, the data is represented in the form of straight lines.
3. Each line in the diagram represents an observation or a class.
4. The height of the line denotes the magnitude of the observation / class.
5. The distance between the lines is kept uniform.
6. Advantages of line diagram: quick and simple method, comparison become easy.

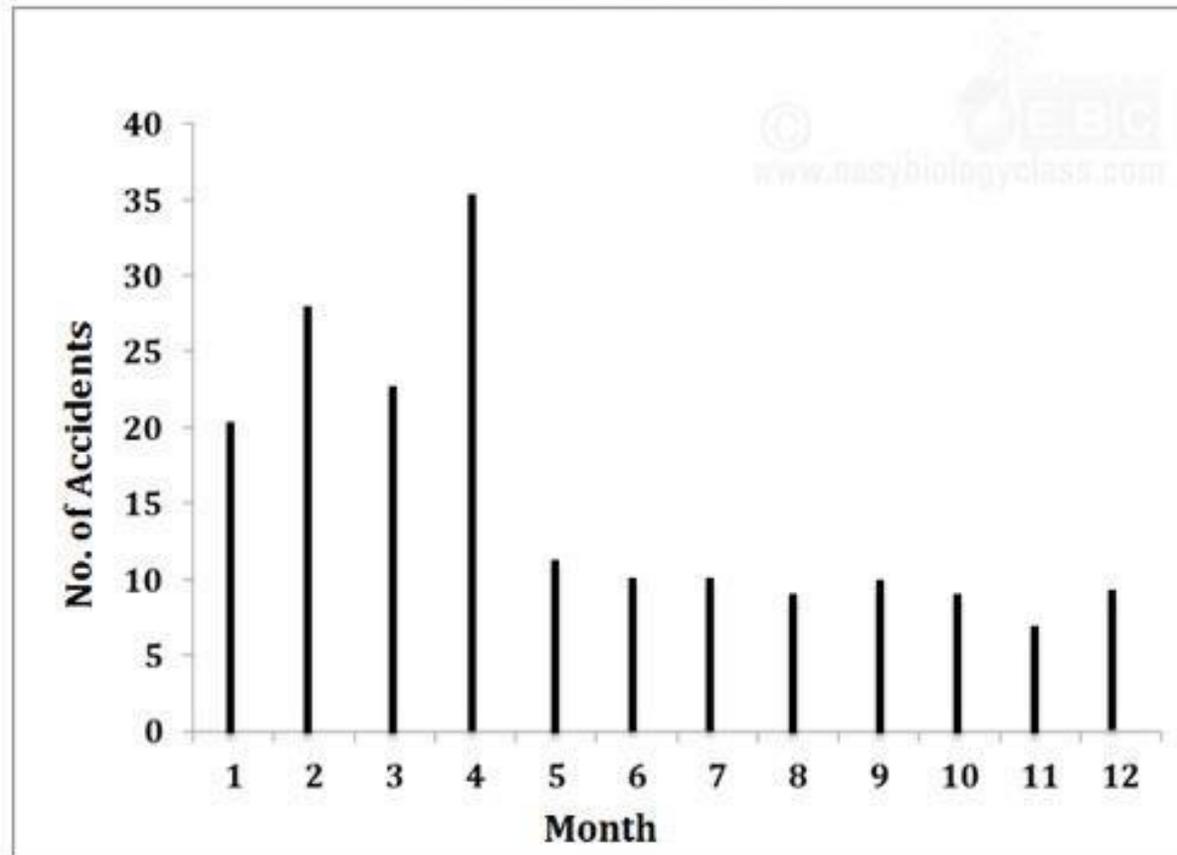
Example: A study on the number of accidents in the year 2015 in a particular area is given below. Draw a line graph to represent the data.

Month	1	2	3	4	5	6	7	8	9	10	11	12
No. of Accidents	21	27	23	34	12	11	9	8	9	8	7	8

Solution

Month	1	2	3	4	5	6	7	8	9	10	11	12
No. of Accidents	21	27	23	34	12	11	9	8	9	8	7	8

LINE GRAPH



(2). Bar Diagram

1. Bar diagram is also called as bar chart
2. A common and simple method of graphical representation of data.
3. Bar diagram is a chart that presents grouped data with rectangular bars.
4. Each rectangular bar represents a class.
5. Height of the bar is proportional to the magnitude of the item in the class
6. Bars are drawn vertically or horizontally with equal spacing between them.
7. The width of the bars and the space between them are kept constant.
8. The vertical bar diagram is also called as column bar chart.
9. In a vertical bar diagram, the independent variables are shown on the X axis, while the dependent variables are shown on the Y axis.
10. Items are to be compared with respect to a single characteristic.

11. Simple bar diagram may be vertical or horizontal.

Bar diagram is further divided into FOUR types:

(a). Simple bar diagram

(b). Multiple bar diagram

(c). Subdivided bar diagram

(d). Percentage bar diagram

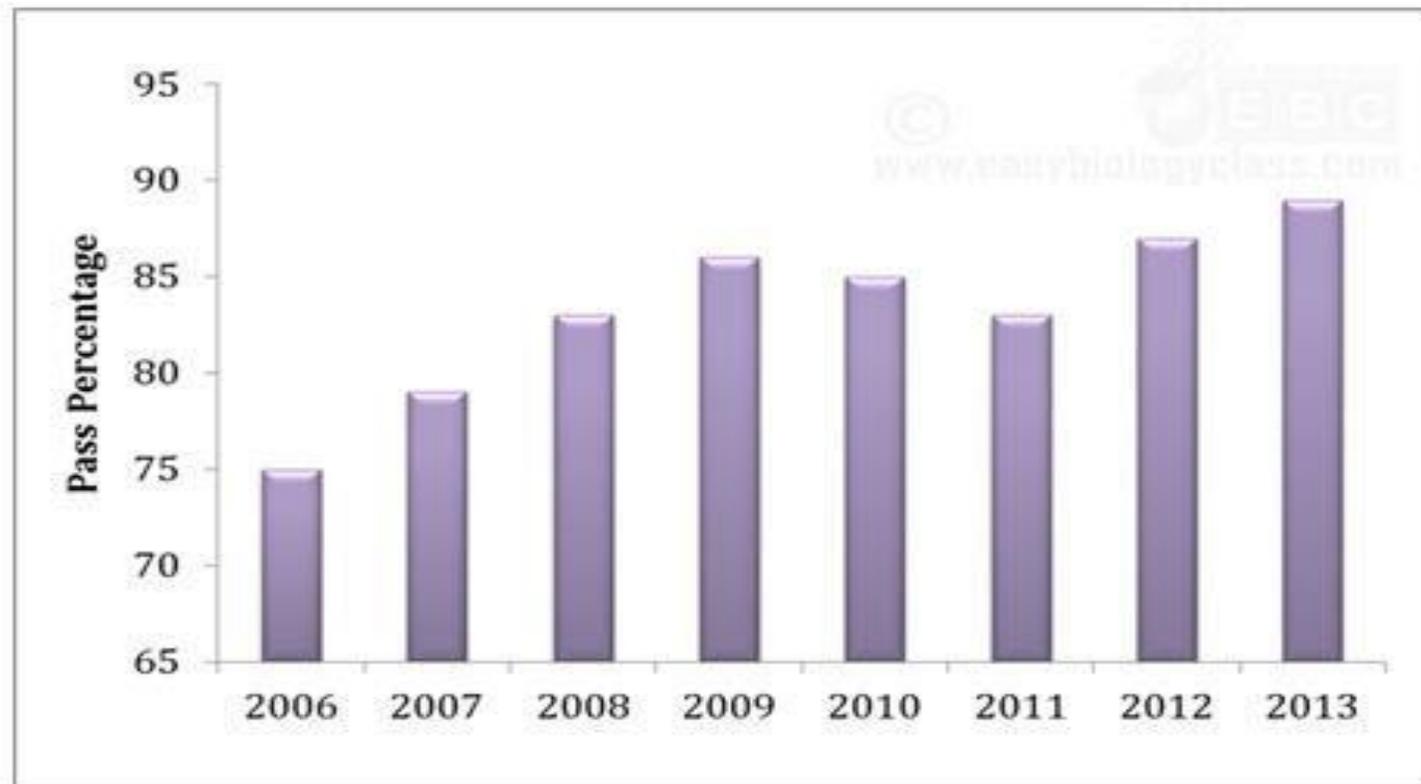
Example: Draw a simple bar diagram using the following data.

<i>Year</i>	2006	2007	2008	2009	2010	2011	2012	2013
<i>Pass Percentage (%)</i>	75	79	83	86	85	83	87	89

Solution

<i>Year</i>	2006	2007	2008	2009	2010	2011	2012	2013
<i>Pass Percentage (%)</i>	75	79	83	86	85	83	87	89

SIMPLE VERTICAL BAR DIAGRAM



(b). Multiple bar diagram

1. Contain two or more bars arranged side by side.
2. Allow comparison of multiple sets of variables comparison.
3. Different colors or shades are used to distinguish different bars in a single set

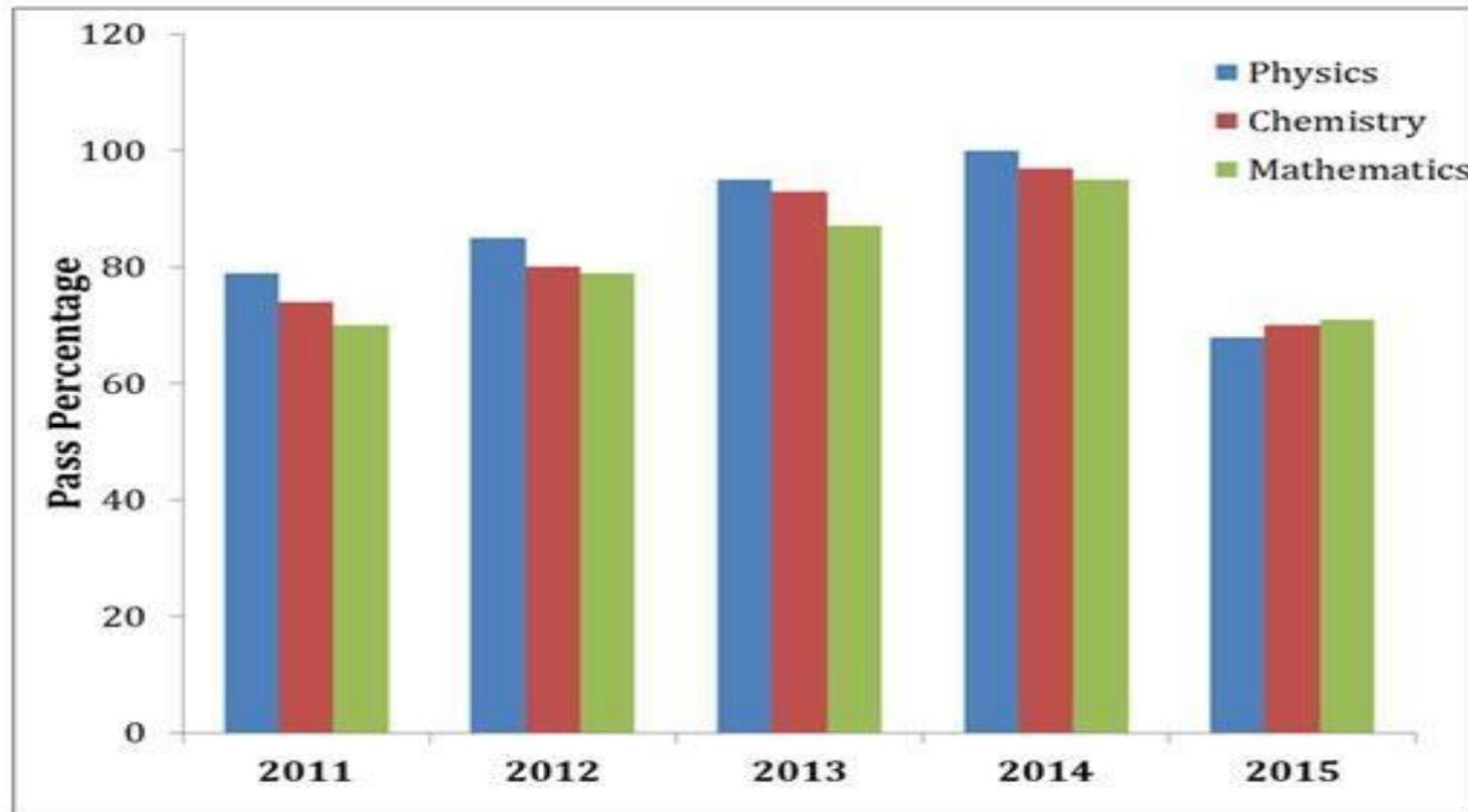
Example: Draw a bar diagram using the following data showing the pass percentage of different subjects in five years

	2011	2012	2013	2014	2015
Physics	79	85	95	100	68
Chemistry	74	80	93	97	70
Mathematics	70	79	87	95	71

Solution

	2011	2012	2013	2014	2015
Physics	79	85	95	100	68
Chemistry	74	80	93	97	70
Mathematics	70	79	87	95	71

MULTIPLE VERTICAL BAR DIAGRAM



(c). Subdivided bar diagram

1. Also called component bar diagram
2. The individual bar is subdivided into various parts or compartments.
3. The size of various compartments is proportional to the magnitude of the variables.
4. Different colours or shades are used to distinguish the compartments of the bar.

5. The distance between the bar and the width of the bar is kept constant.

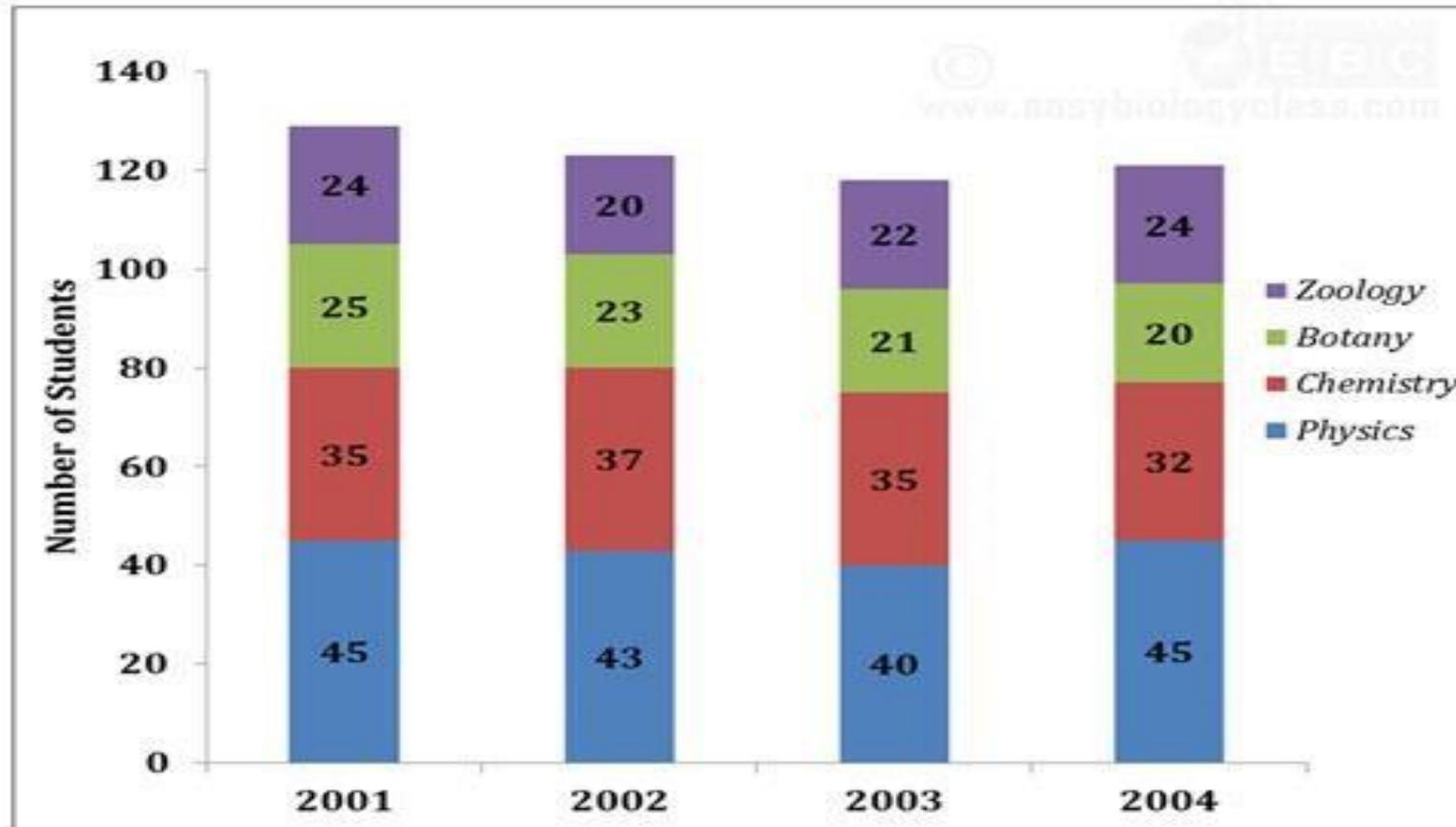
Example: Number of science graduate students in a college is given below. Draw a subdivided bar diagram using the following data.

Subject	2001	2002	2003	2004
Physics	45	43	40	45
Chemistry	35	37	35	32
Botany	25	23	21	20
Zoology	24	20	22	24

Solution

Subject	2001	2002	2003	2004
Physics	45	43	40	45
Chemistry	35	37	35	32
Botany	25	23	21	20
Zoology	24	20	22	24

SUBDIVIDED BAR DIAGRAM



(d). Percentage bar diagram

1. Percentage bar diagram is a diagram which exhibits a simple analysis of statistical data in terms of percentage.
2. The length of all bars is kept constant (100%).
3. Each bar consists of several compartments.

4. The size of each compartment of a bar corresponds to the percentage of that component with respect to the total.

Example: Draw a percentage bar diagram using the following data.

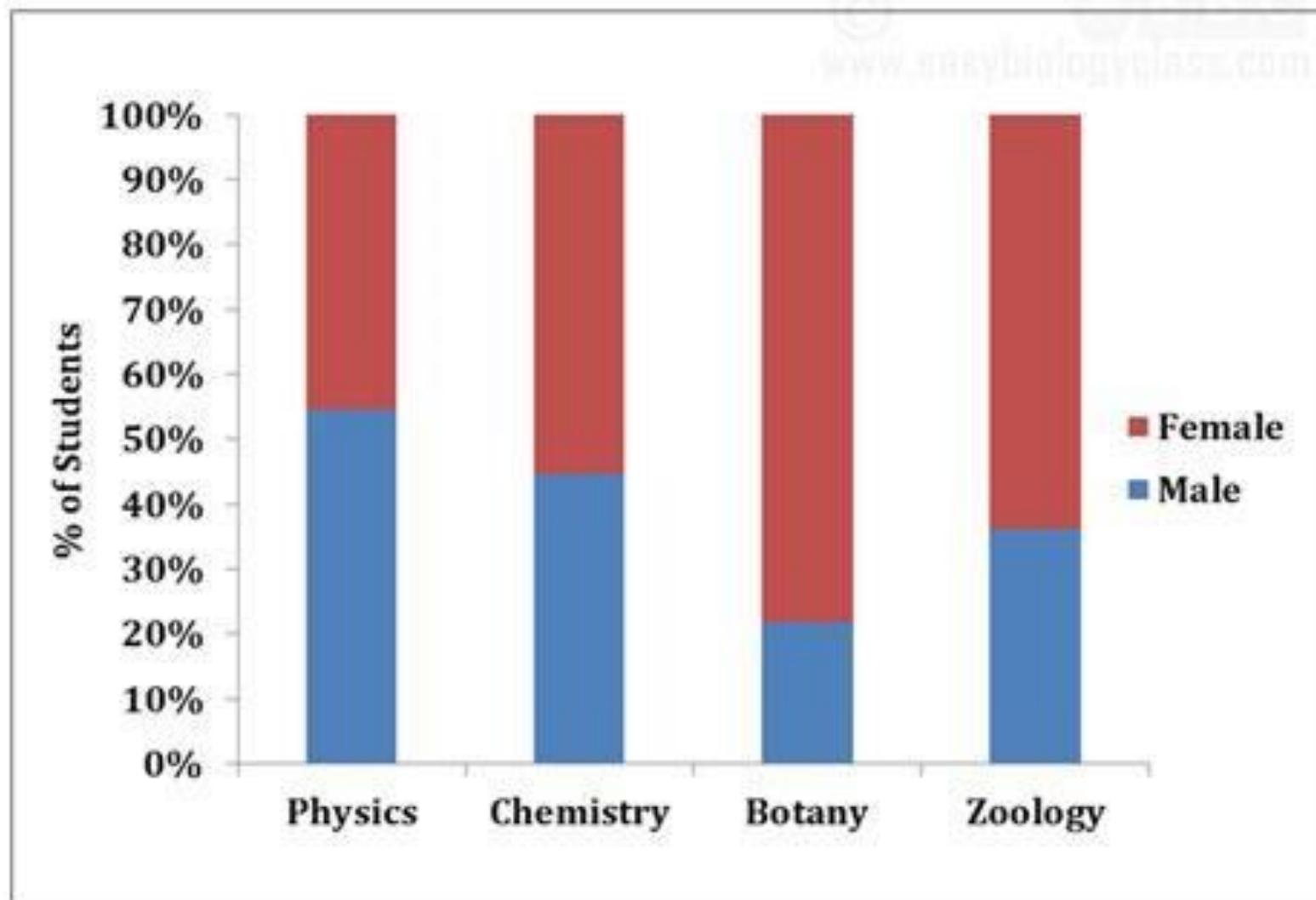
	Physics	Chemistry	Botany	Zoology
Male Students	24	21	07	13
Female Students	20	26	25	23
Total	44	47	32	36

Solution: Conversion of absolute values into percentage.

	Physics	Chemistry	Botany	Zoology
Male Students	24	21	07	13
Female Students	20	26	25	23
Total	44	47	32	36

	Physics	Chemistry	Botany	Zoology
Male Students	$(24/44) \times 100$ = 54.55%	$(21/47) \times 100$ = 44.68%	$(07/32) \times 100$ = 21.87%	$(13/36) \times 100$ = 36.11%
Female Students	$(20/44) \times 100$ = 45.45%	$(26/47) \times 100$ = 55.32%	$(25/32) \times 100$ = 78.13%	$(23/36) \times 100$ = 63.89%

PERCENTAGE BAR DIAGRAM



(2). Histogram

1. Histogram is used in the graphical representation of frequency distribution.
2. Here each class of the frequency distribution is represented as columns.
3. The height of the column corresponds to the magnitude of the frequency.
4. A histogram quickly tells how many items are there in each numerical category.
5. The histogram resembles a bar diagram (but with a difference).

6. In the histogram, the columns representing each class are in close contact and there is no space between them.
7. The absence of inter-bar space denotes the continuity of classes in the histogram.

Construction of a Histogram

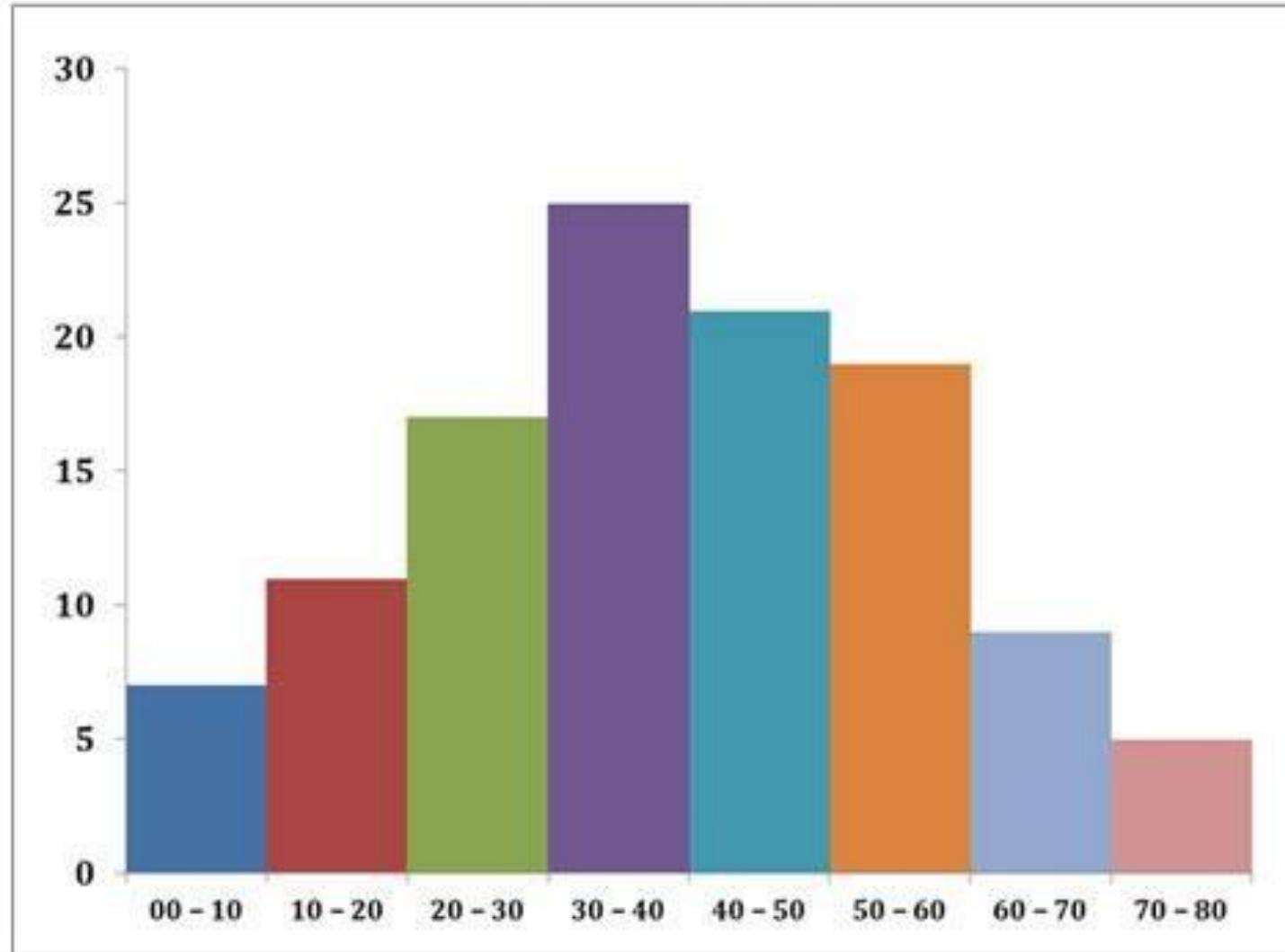
1. The class intervals are taken on the X axis.
2. Corresponding frequencies are taken on the Y axis.
3. Class intervals used are usually of equal width.
4. The frequency is proportional to the area and height of the bar.

Example: Construct a histogram using the following data

Class	00 - 10	10 - 20	20 - 30	30 - 40	40 - 50	50 - 60	60 - 70	70 - 80
Frequency	7	11	17	25	21	19	9	5

Solution:

HISTOGRAM



Importance of histogram

1. The area of blocks in the histogram clearly shows the frequency of each class.
2. Histogram shows the spread of observations (uniformly spread or randomly spread or showing central tendency).

3. Provide information about skewness or symmetry of data.

(3). Frequency Polygon

1. The Frequency Polygon is a curve representing a frequency distribution.
2. In frequency polygon, the mid values of each class are first obtained.
3. In a graph paper, the frequency of each class is plotted against the mid value of class (on the X axis).
4. Then these points are then joined by a straight line.
5. This straight line is extended in both directions to meet on the X axis.

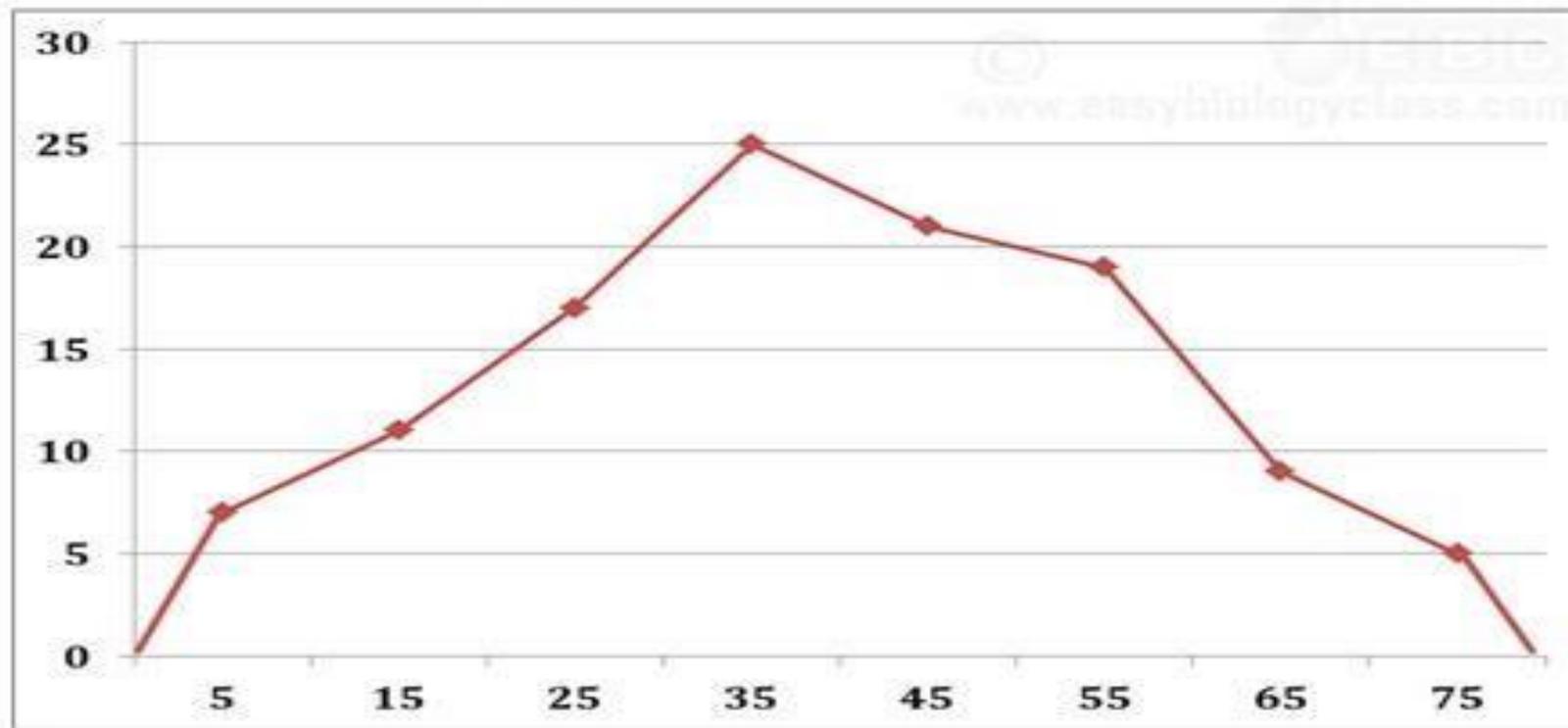
6. The first point is joined to the lower limit of the first class and the last point is joined to the upper limit of last class. Thus, the frequency polygon is a closed graph.
7. The graph now obtained is called Frequency polygon.

Example: Construct a Frequency Polygon using the following data

Class	00 – 10	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60	60 – 70	70 – 80
Frequency	7	11	17	25	21	19	9	5

Class	00 - 10	10 - 20	20 - 30	30 - 40	40 - 50	50 - 60	60 - 70	70 - 80
Mid-value	5	15	25	35	45	55	65	75
Frequency	7	11	17	25	21	19	9	5

FREQUENCY POLYGON



(4). Frequency curve

1. The construction of a frequency curve is similar to that of a frequency polygon.
2. Here the points of frequencies in the graph are connected by a free-hand smooth curve.
3. Thus, frequency curve is also called as 'smoothed frequency curve'.
4. The frequency curve also begins and ends in base line (X axis).

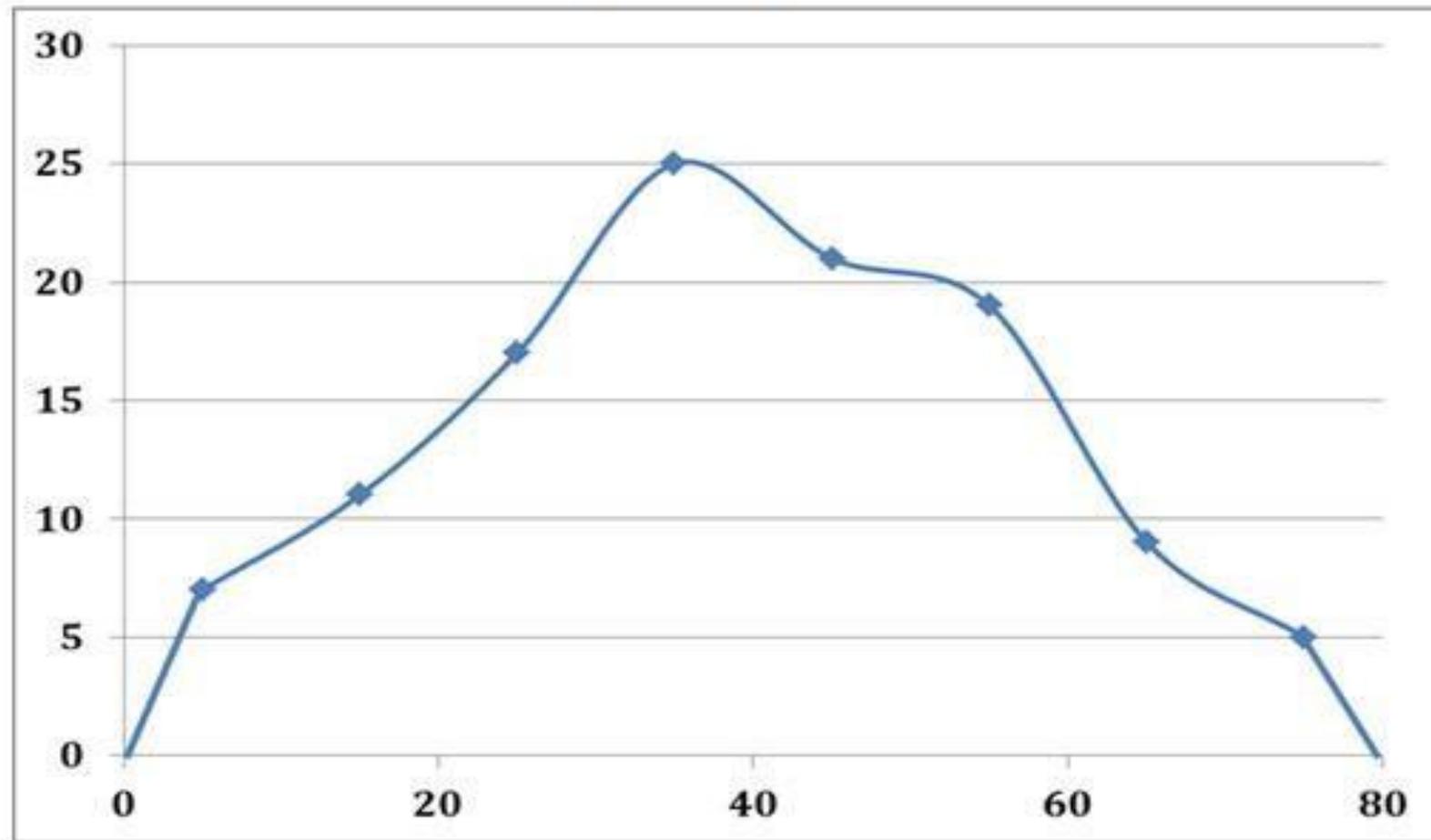
Advantage: More smooth appearance of data than frequency polygon.

Example: Construct a Frequency curve using the following data

Class	00 – 10	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60	60 – 70	70 – 80
Frequency	7	11	17	25	21	19	9	5

Class	00 - 10	10 - 20	20 - 30	30 - 40	40 - 50	50 - 60	60 - 70	70 - 80
Mid-value	5	15	25	35	45	55	65	75
Frequency	7	11	17	25	21	19	9	5

FREQUENCY CURVE



(5). Ogive

1. The ogive is a cumulative frequency curve.
2. It is a free-hand graph showing the curve of a cumulative frequency.
3. The ogive is constructed by plotting the upper class limit on the X axis and the corresponding cumulative frequency on the Y axis.
4. Ogive is best used when the total frequency at any given time is to be displayed.
5. There are TWO types of ogives:
 - (1). Less than ogive
 - (2). Greater than ogive (more than ogive)
 - (1). Less than Ogive

Less than ogive is the graph of the less than cumulative frequency distribution which shows the number of observations LESS THAN the upper class limit.

Construction of an Ogive

1. First we calculate the cumulative frequency (as usual).
2. Then plot the upper class limit of class interval against the value of corresponding cumulative frequency.
3. The points are joined by a free-hand smooth curve.

4. The lower point of the curve is joined to the X-axis at the lower limit of the first class interval.

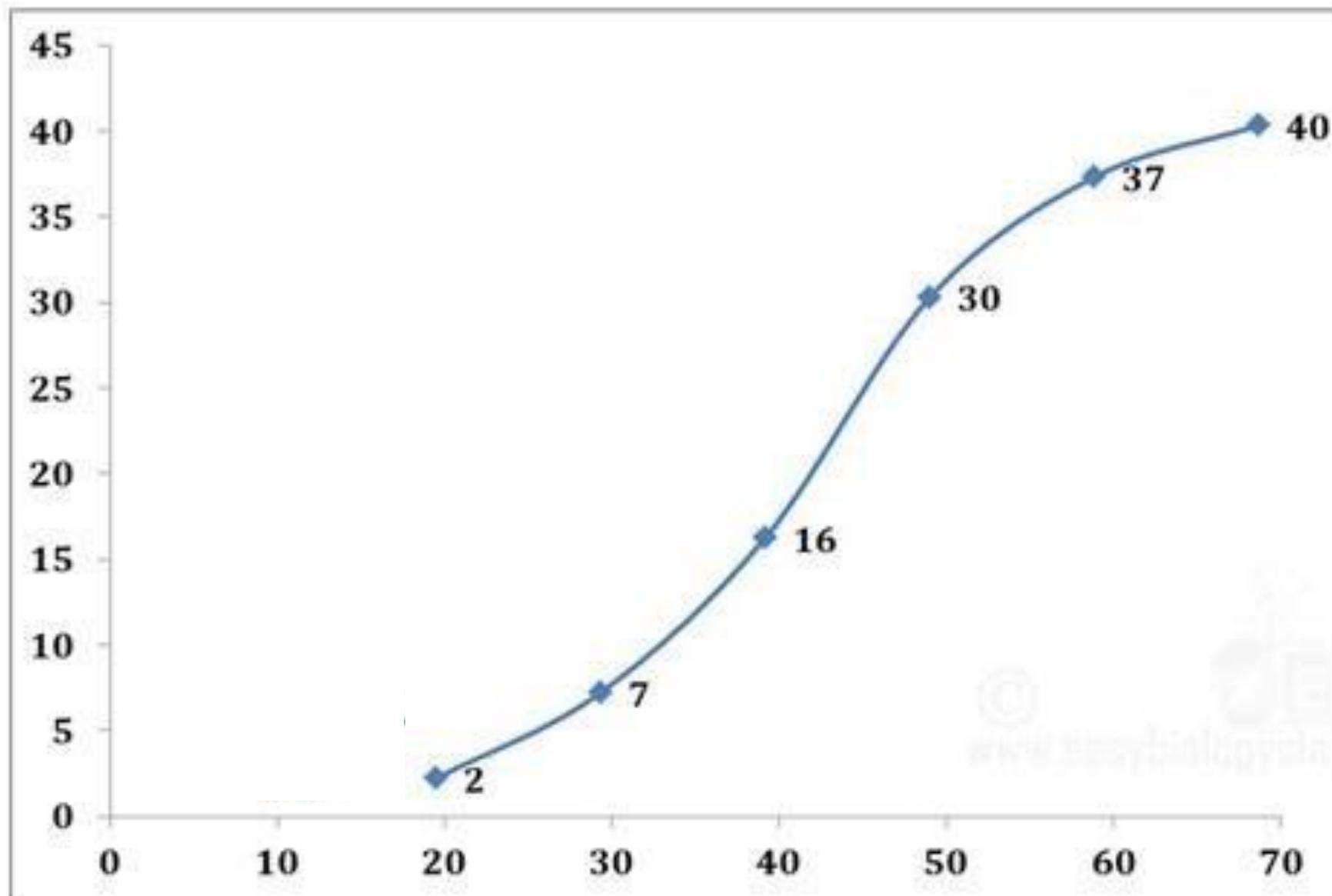
Example: Construct a less than ogive using the following data

Class	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60	60 – 70
Frequency	2	5	9	14	7	3

Solution: Find the upper class limit and cumulative frequency. Then plot the cumulative frequency against the upper class limit.

Class	Frequency	Upper Limit of Class	Cumulative Frequency
10 - 20	2	20	2
20 - 30	5	30	7
30 - 40	9	40	16
40 - 50	14	50	30
50 - 60	7	60	37
60 - 70	3	70	40

LESS THAN OGIVE



(2). Greater than ogive

1. Greater than ogive is the graph of the greater than cumulative frequency distribution which shows the number of observations **GREATER THAN** the lower class limit.
2. First, the frequencies are cumulated from below
3. Then the cumulated frequency is plotted against the lower limit of the corresponding class interval.

4. The uppermost point is joined with the Y-axis at the upper limit of the last class.

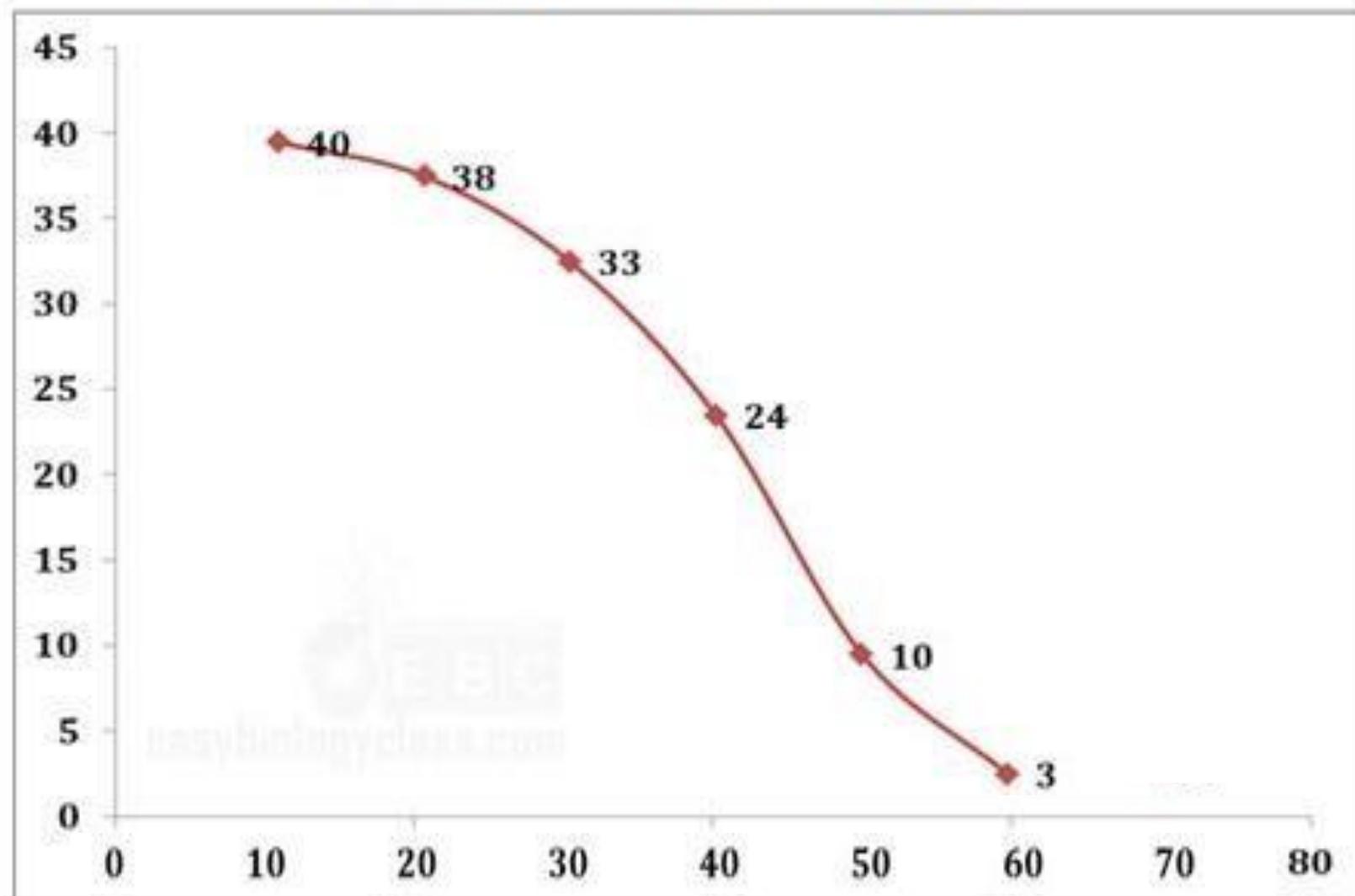
Example: Construct a greater than ogive using the following data

Class	10 - 20	20 - 30	30 - 40	40 - 50	50 - 60	60 - 70
Frequency	2	5	9	14	7	3

Solution: Find the lower class limit and cumulative frequency from below. Then plot the cumulative frequency against the lower class limit .

Class	Frequency	Lower Limit of Class	Cumulative Frequency
10 – 20	2	10	40
20 – 30	5	20	38
30 – 40	9	30	33
40 – 50	14	40	24
50 – 60	7	50	10
60 – 70	3	60	3

GREATER THAN OGIVE

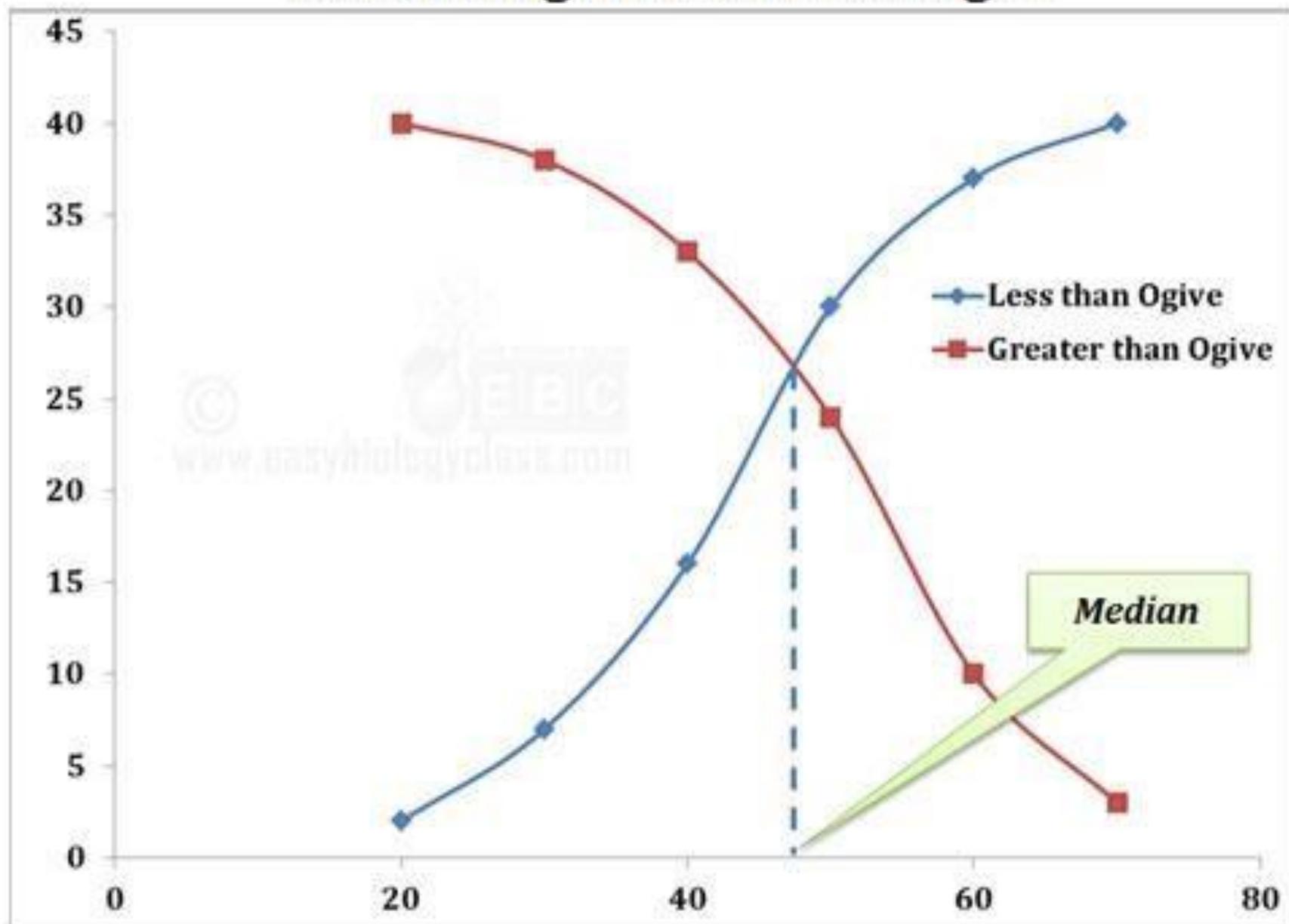


Calculating Median from Ogive

1. Median of the frequency distribution can be calculated from the ogive with a quick method.
2. First, draw a less than ogive and then draw a greater than ogive of the same data on same graph paper.
3. The two curves will intercept at some point.

4. A perpendicular line is drawn from the point of interception to the X -axis.
5. This value on the X axis will be the median of the frequency distribution.

Calculating Median from Ogive



(6). Pie Chart

1. The pie chart is also called as the circle graph or area diagram
2. The pie chart makes the use of sectors in a circle to represent different categories of data.
3. The circle has a total area of 360° which can be divided into component sectors based on the given data.
4. The arc length of each sector is proportional to the frequency of the data.

5. Arc length is proportional to central angle and the area of sector.

How to construct a Pie chart

1. Calculate the angle of each sector (class) corresponding the frequency using the formula

$$\textit{Angle of sector} = \frac{\textit{Value of the category}}{\textit{Total value of all categories}} \times 360$$

2. Draw a circle of any radius and mark the center.
3. Mark the angles of each sector accurately
4. Sectors can be then labeled or shaded to denote different classes.

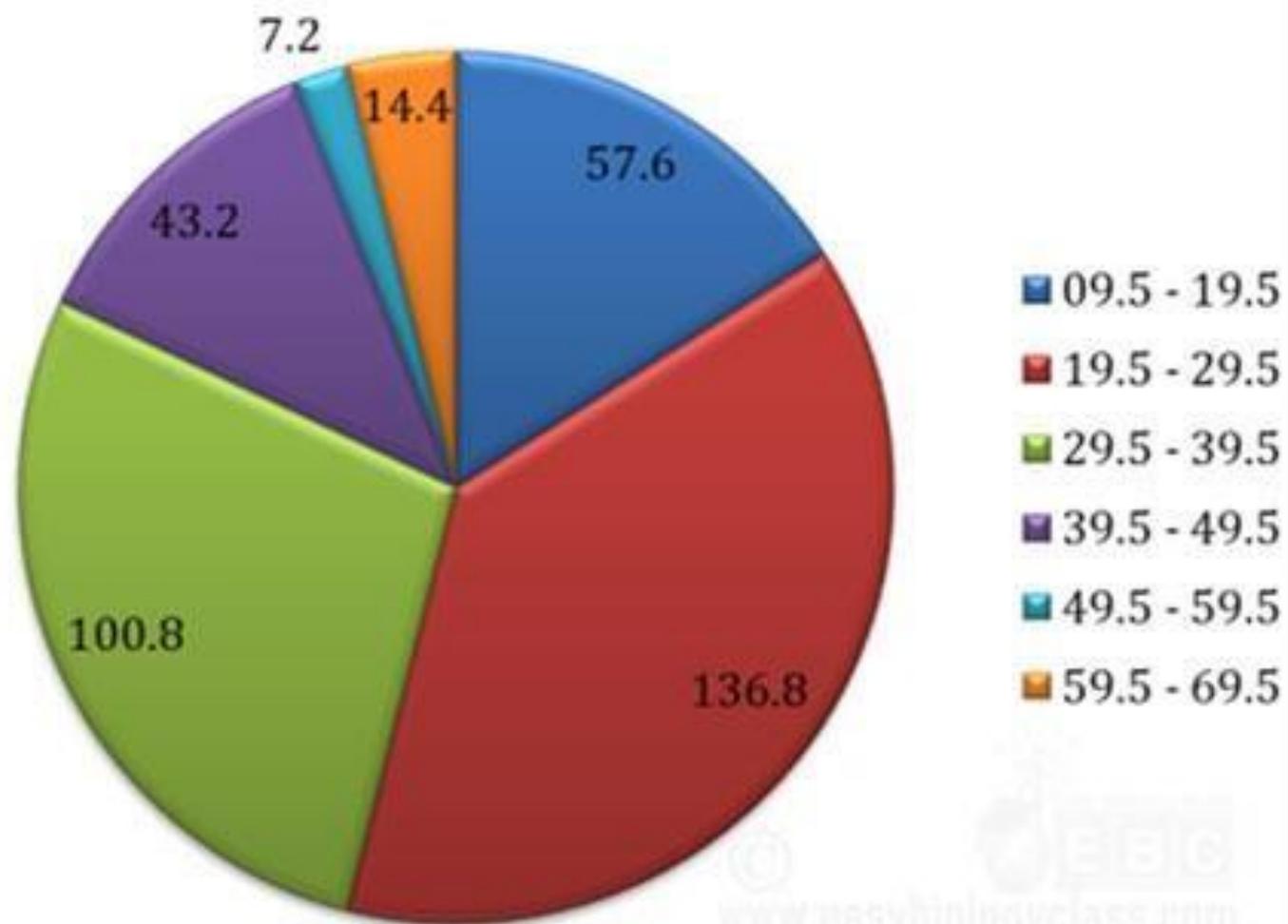
Example: Construct a pie diagram using the following data

Class	Frequency
09.5 - 19.5	8
19.5 - 29.5	19
29.5 - 39.5	14
39.5 - 49.5	6
49.5 - 59.5	1
59.5 - 69.5	2
	50

Solution: First calculate the angle of each sector using the above equation.

Class	Frequency	Angle of Sector
09.5 - 19.5	8	57.6
19.5 - 29.5	19	136.8
29.5 - 39.5	14	100.8
39.5 - 49.5	6	43.2
49.5 - 59.5	1	7.2
59.5 - 69.5	2	14.4
	50	360

PIE DIAGRAM



The pie diagram can also be drawn using relative frequency using the formula

$$\textit{Angle of sector} = \frac{\mathbf{360 \times Relative Frequency}}{\mathbf{100}}$$

Relative frequency: The frequency expressed as a fraction of the total frequency and the fraction of the obtained frequency.

The relative frequency is expressed as the percentage.

Measures of Central Tendency

Contents

Concept of Mean or Average

Sigma notation

Arithmetic Mean (A.M.)

Weighted A.M.

A.M. from grouped data

Missing frequency problem

Geometric Mean (G.M.)

Harmonic Mean (H.M.)

Median

Mode

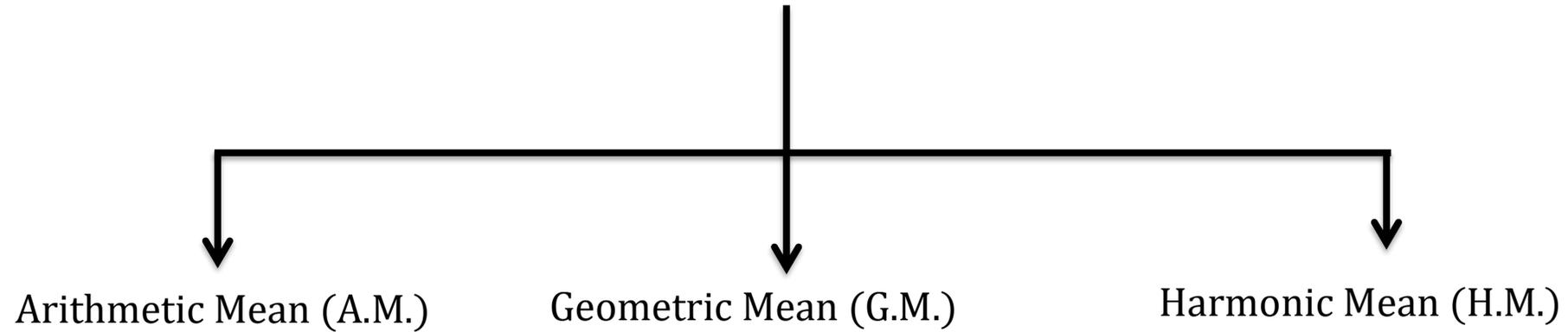
Sigma notation Σ

The sigma notation is a shorthand notation used to sum up a large number of terms.

$$\sum x = \sum_{i=1}^n x_i = x_1 + x_2 + \cdots + x_n$$

One uses this notation because it is more convenient to write the sum in this fashion.

Three types of Mean



Definition of the A.M.

- Given a sample of n data points, $x_1, x_2, x_3, \dots, x_n$, the formula for the arithmetic mean is given below.

$$\text{Arithmetic Mean} = \frac{\textit{Sum of the data points}}{\textit{Total number of data points}} = \frac{\sum x}{n}$$

Find the Arithmetic Mean

- Find the A.M. of the numbers 2, 5, 9, 11 and 13.

- Answer: A.M. is,

$$\begin{aligned} & \frac{2 + 5 + 9 + 11 + 13}{5} \\ &= \frac{40}{5} \\ &= 8 \end{aligned}$$

Weighted A.M.

- Values of the variates x_1, x_2, \dots, x_n are taken f_1, f_2, \dots, f_n times respectively.

- Weighted Arithmetic Mean
$$\frac{\sum fx}{\sum f} = \frac{\sum_{i=1}^n f_i x_i}{\sum f}$$

Example: Find the average income and the total income from the following table

Daily income (Rs.)	2	5	9	11	13	Total
Number of workers	2	1	4	2	1	10

Variate (x)	Frequency (f)	fx
(income in Rs.)	(no. of workers)	
2	2	4
5	1	5
9	4	36
11	2	22
13	1	13
Total	$n = \sum f = 10$	$\sum fx = 80$

$$\text{A.M.} = \frac{80}{10} = 8$$

Calculation of mean from grouped data

Example : Calculate the mean weight from the following table

Weight (lbs.)	95-105	105-115	115-125	125-135	total
No. of students	20	26	38	16	100

weight	No. of students (f)	Mid value (x)	fx		
95-105	20	100	2000		
105-115	26	110	2860		
115-125	38	120	4560		
125-135	16	130	2080		
Total	100		11500		

$$\text{A.M.} = \frac{\sum fx}{\sum f} = \frac{11500}{100} = 115$$

Finding missing frequency problem

Example: A.M. of the following is 67.45. Find the value of f_3

Height: 60-62 63-65 66-68 69-71 72-44

Frequency: 15 54 f_3 81 24

Height	frequency (f)	mid value (x)	fx
60-62	15	61	915
63-65	54	64	3456
66-68	f_3	67	$67 f_3$
69-71	81	70	5670
72-44	24	73	1752
total	$\Sigma f = 174 + f_3$		$\Sigma fx = 11793 + 67 f_3$

$$\text{A.M.} = \frac{\Sigma fx}{\Sigma f} \Rightarrow 67.45 = \frac{11793 + 67 f_3}{174 + f_3} \Rightarrow f_3 = 126$$

Geometric Mean (G.M.)

Suppose given variates are x_1, x_2, \dots, x_n

$$\text{Then G.M. is } \sqrt[n]{x_1 \times x_2 \times \dots \times x_n}$$

Example: Find the G.M. of 111,171,191,212.

$$\text{G.M.} = \sqrt[4]{111 \times 171 \times 191 \times 212} = 166.5$$

Harmonic Mean (H.M.)

$$\text{H.M.} = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n}}$$

Example : Find the H.M. of 3, 6, 12, 15

$$\text{H.M.} = \frac{4}{\frac{1}{3} + \frac{1}{6} + \frac{1}{12} + \frac{1}{15}} = 6.15$$

Relations between A.M., G.M. and H.M.

1. $A.M. \geq G.M. \geq H.M.$

Median :

If a set of observations is arranged in order of magnitude (ascending or descending) then the middle most value gives median.

(A) For simple frequency distribution:

step 1: arrange in order of magnitude (ascending or descending)

step 2:

median = value of $\frac{n+1}{2}$ th item

where n is the total frequency.

(B) For continuous series distribution:

$$\begin{aligned}\text{median} &= \text{value of } \frac{n}{2}\text{th item} \\ &= l_1 + \frac{\frac{n}{2} - c}{f_m} (l_2 - l_1)\end{aligned}$$

where, $l_1 =$ lower limit of the median class

$l_2 =$ upper limit of the median class

$f_m =$ frequency of the median class

$c =$ cumulative frequency of the class preceding median class

Mode

Mode is the value of the variate which occurs most frequently.
It represents the most frequent value of the series.

Mode for class interval type
distribution

$$= l_1 + \frac{f_0 - f_1}{2f_0 - f_1 - f_2} (l_2 - l_1)$$

where, l_1 = lower limit of the median class

l_2 = upper limit of the median class

f_m = frequency of the median class

c = cumulative frequency of the class preceding median class

Chapter - 3

Measures of Dispersion

- Types

<p>A. Absolute Measures</p> <p>Range Quartile Deviation (Q.D.) Mean Deviation (M.D.) Standard Deviation (S.D.)</p>	<p>B. Relative Measures</p> <p>Coefficient of Range Coefficient of Quartile Deviation Coefficient of Mean Deviation Coefficient of Standard Deviation</p>
--	---

Range

For Simple data distribution

$$\text{Range} = \text{Maximum value} - \text{Minimum value} = L - S$$

For Grouped frequency distribution

$$\text{Range} = \text{Upper limit of highest C.I.} - \text{Lower limit of lowest C.I.}$$

$$\text{Coefficient of Range} = \left(\frac{L-S}{L+S} \right) = \left(\frac{L-S}{L+S} \right) \times 100\%$$

Quartile Deviation

$$\text{Quartile Deviation} = \frac{1}{2}(Q_3 - Q_1)$$

$$\text{Coefficient of Quartile Deviation} = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

Mean Deviation

$$\text{Mean Deviation} = \frac{\sum f |x_i - \bar{x}|}{\sum f}$$

where \bar{x} is the mean about which M.D. is required.

$$\text{Coefficient of Mean Deviation} = \frac{M.D.}{\bar{x}}$$

Standard Deviation

$$S.D. = \sqrt{\frac{\sum f(x - \bar{x})^2}{\sum f}} = \sqrt{\frac{\sum fx^2}{\sum f} - \left(\frac{\sum fx}{\sum f}\right)^2}$$

where \bar{x} = A.M. of the distribution

$$\text{Coefficient of Standard Deviation} = \frac{S.D.}{\bar{x}}$$

Chapter - 4

MOMENTS, SKEWNESS AND KURTOSIS

Beyond the measures of central tendency and dispersion explained earlier, there are measures that further describe the characteristics of a distribution. Some of them are discussed here.

Moments

Moments are a set of statistical parameters to measure a distribution.

The formula for calculating moments is as follows:

$$\text{1st moment} = \mu_1 = \frac{\sum f(x-\bar{x})}{\sum f}$$

$$\text{2nd moment} = \mu_2 = \frac{\sum f(x-\bar{x})^2}{\sum f}$$

⋮

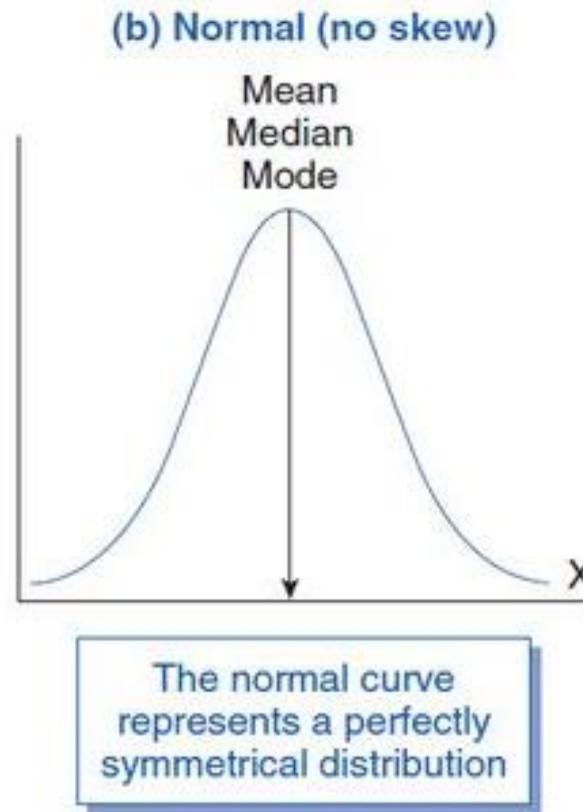
$$\text{n-th moment} = \mu_n = \frac{\sum f(x-\bar{x})^n}{\sum f}$$

Skewness

The term 'skewness' refers to lack of symmetry or departure from symmetry, e.g., when a distribution is not symmetrical (or is asymmetrical) it is called a skewed distribution.

The measures of skewness indicate the difference between the manner in which the observations are distributed in a particular distribution compared with a symmetrical (or normal) distribution.

In a symmetrical distribution, the values of mean, median and mode are alike.

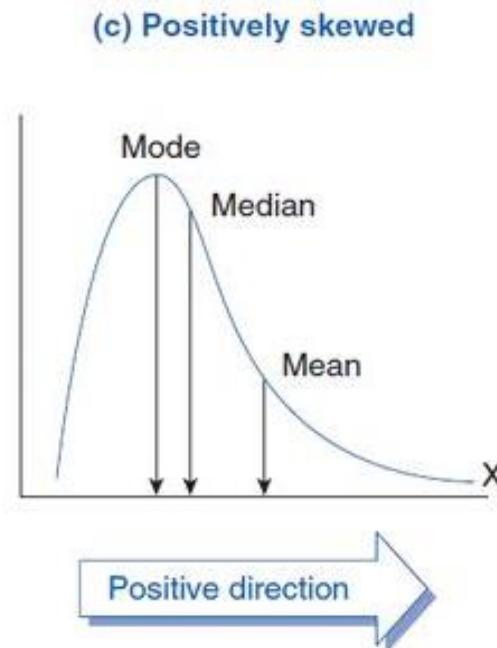


If the value of mean is greater than the mode, skewness is said to be positive.

In a positively skewed distribution, mean is greater than the mode and the median lies somewhere in between mean and mode.

A positively skewed distribution contains some values that are much larger than most other observations.

A distribution is positively skewed when the long tail is on the positive side of the peak.



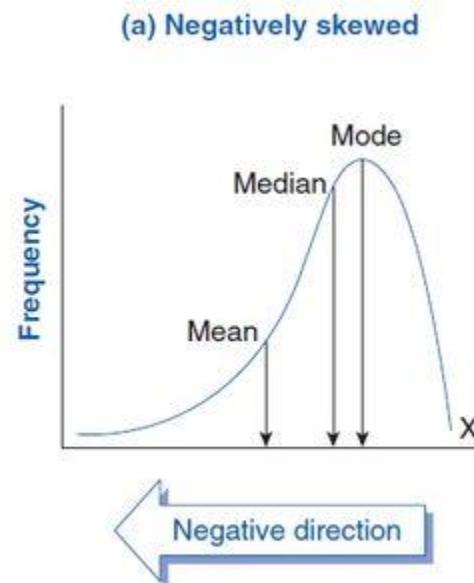
If the value of mode is greater than mean, skewness is said to be negative.

The following diagrams could clarify the meaning of skewness.

In a negatively skewed distribution, mode is greater than the mean and the median lies in between mean and mode.

The mean is pulled towards the low-valued item (that is, to the left). A negatively skewed distribution contains some values that are much smaller than most observations.

A distribution is negatively skewed when the long tail is on the negative side of the peak.



Generally, If Mean $>$ Mode, the skewness is positive.

If Mean $<$ Mode, the skewness is negative.

If Mean = Mode, the skewness is zero.

Skewness is measured in the following ways:

$$\text{Karl Pearson's Coefficient of Skewness} = \frac{\text{Mean} - \text{Mode}}{\text{Standard Deviation}} \text{ or } \frac{3(\text{Mean} - \text{Median})}{\text{Standard Deviation}}$$

$$\text{Moment based measure of skewness} = \beta_1 = \frac{\mu_3^2}{\mu_2^3}$$

$$\text{Pearson's coefficient of skewness} = \gamma_1 = \sqrt{\beta_1}$$

Kurtosis

Kurtosis refers to the degree of peakedness of a frequency curve.

It tells how tall and sharp the central peak is, relative to a standard bell curve of a distribution.

Kurtosis can be described in the following ways:

- Platykurtic -

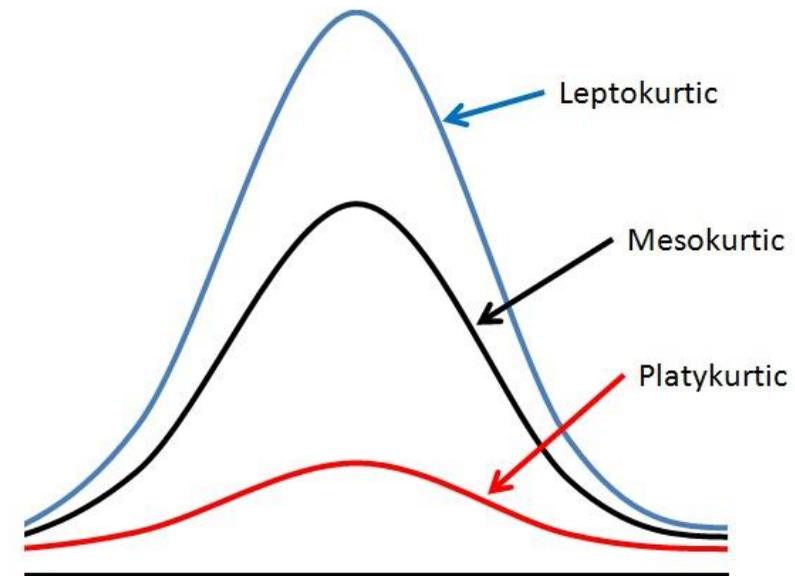
When the kurtosis < 0 , the frequencies throughout the curve are closer to be equal (i.e., the curve is more flat and wide)

- Leptokurtic -

When the kurtosis > 0 , there are high frequencies in only a small part of the curve (i.e, the curve is more peaked)

- Mesokurtic -

When the kurtosis $= 0$



Kurtosis is measured in the following ways:

$$\text{Moment based Measure of kurtosis} = \beta_2 = \frac{\mu_4}{\mu_2^2}$$

$$\text{Coefficient of kurtosis} = \gamma_2 = \beta_2 - 3$$

Chapter 5 - INTERPOLATION

- **Introduction**
- **Difference operator**
- **Shift operator**
- **Newton's Forward Difference Interpolation Formula**
- **Newton's Backward Difference Interpolation Formula**
- **Lagrange's Interpolation**

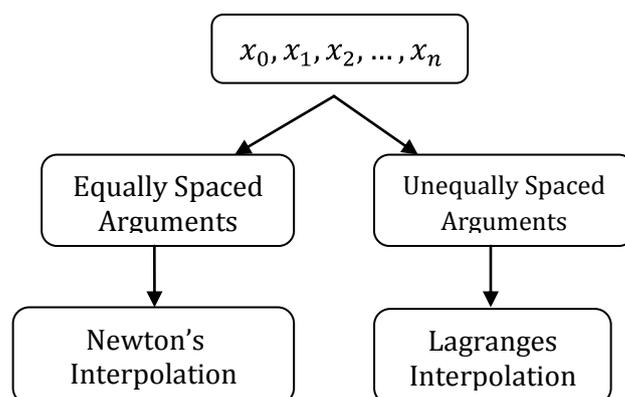
INTERPOLATION

The process of finding the curve passing through the points $(x_0, y_0), (x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ is called as Interpolation and the curve obtained is called as Interpolating curve.

Interpolating polynomial passing through the given set of points is unique.

Let $x_0, x_1, x_2, \dots, x_n$ be given set of observations and $y = f(x)$ be the given function, then the method to find $f(x_m) \forall x_0 \leq x_m \leq x_n$ is called as an Interpolation.

If x_m is not in the range of x_0 and x_n , then the method to find $f(x_m)$ is called as Extrapolation.



The Interpolation depends upon finite difference concept.

If $x_0, x_1, x_2, \dots, x_n$ be given set of observations and let $y_0 = f(x_0), y_1 = f(x_1), \dots, y_n = f(x_n)$ be their corresponding values for the curve $y = f(x)$, then $y_1 - y_0, y_2 - y_1, \dots, y_n - y_{n-1}$ is called as finite difference.

Difference operator

Let us consider $x_0, x_1, x_2, \dots, x_n$ be given set of observations and let $y_0, y_1, y_2, \dots, y_n$ are corresponding values of the curve $y = f(x)$, then the Forward difference operator is denoted by Δ and is defined as $\Delta y_0 = y_1 - y_0, \Delta y_1 = y_2 - y_1, \dots, \Delta y_{n-1} = y_n - y_{n-1}$.

In this case $\Delta y_0, \Delta y_1, \dots, \Delta y_n$ are called as First Forward differences of y .

The difference of first forward differences will give us Second forward differences and it is denoted by Δ^2 and is defined as $\Delta^2 y_0 = \Delta(\Delta y_0)$

$$= \Delta(y_1 - y_0) = \Delta y_1 - \Delta y_0$$

$$= (y_2 - y_1) - (y_1 - y_0)$$

$$= y_2 - 2y_1 - y_0$$

Similarly, the difference of second forward differences will give us third forward difference and it is denoted by Δ^3 .

difference table

x	$y = f(x)$	First Forward differences Δy	Second Forward differences $\Delta^2 y$	Third Forward differences $\Delta^3 y$	Fourth differences $\Delta^4 y$
x_0	y_0	$\Delta y_0 = y_1 - y_0$			
x_1	y_1		$\Delta^2 y_0 = \Delta y_1 - \Delta y_0$		
x_2	y_2	$\Delta y_1 = y_2 - y_1$		$\Delta^3 y_0 = \Delta^2 y_1 - \Delta^2 y_0$	
x_3	y_3		$\Delta^2 y_1 = \Delta y_2 - \Delta y_1$		$\Delta^4 y_0 = \Delta^3 y_1 - \Delta^3 y_0$
.	.			$\Delta^3 y_1 = \Delta^2 y_3 - \Delta^2 y_2$	
.	.	.	$\Delta^2 y_2 = \Delta y_3 - \Delta y_2$		
.	.	.			
x_{n-1}	y_{n-1}				
x_n	y_n	$\Delta y_{n-1} = y_n - y_{n-1}$			

Note: If h is common difference in the values of x and $y = f(x)$ be the given function then $\Delta f(x) = f(x + h) - f(x)$.

Shift Operator: The Shift operator E is defined by the equation $Ey_r = y_{r+1}$

Similarly, $E^n y_r = y_{n+r}$

(Or)

Let h is the common difference in the values of x and $y = f(x)$ be the given function, then the shift operator is denoted by E and is defined as $Ef(x) = f(x + h)$

Inverse Operator: The Inverse Operator E^{-1} is defined as $E^{-1}y_r = y_{r-1}$

In general, $E^{-n}y_r = y_{r-n}$

Properties

1) Prove that $E = 1 + \Delta$

Sol: Consider R.H.S: $(1 + \Delta)y_n = y_n + \Delta y_n$

$$= y_n + (y_{n+1} - y_n)$$

$$= y_{n+1}$$

$$= E^1 y_n \quad (\because E^n y_r = y_{n+r})$$

$$\therefore E = 1 + \Delta$$

Newton's Forward Interpolation Formula

Statement: If $x_0, x_1, x_2, \dots, x_n$ are given set of observations with common difference h and let $y_0, y_1, y_2, \dots, y_n$ are their corresponding values, where $y = f(x)$ be the given function then

$$f(x) = y_0 + p \Delta y_0 + \frac{p(p-1)}{2!} \Delta^2 y_0 + \frac{p(p-1)(p-2)}{3!} \Delta^3 y_0 + \dots + \frac{p(p-1)(p-2)\dots(p-(n-1))}{n!} \Delta^n y_0$$

where $p = \frac{x-x_0}{h}$

Newton's Backward Interpolation Formula

Statement: If $x_0, x_1, x_2, \dots, x_n$ are given set of observations with common difference h and let $y_0, y_1, y_2, \dots, y_n$ are their corresponding values, where $y = f(x)$ be the given function then

$$f(x) = y_n + p \nabla y_n + \frac{p(p+1)}{2!} \nabla^2 y_n + \frac{p(p+1)(p+2)}{3!} \nabla^3 y_n + \dots + \frac{p(p+1)(p+2)\dots(p+(n-1))}{n!} \nabla^n y_n$$

where $p = \frac{x-x_n}{h}$

Lagrange's Interpolation Formula

Statement: If $x_0, x_1, x_2, \dots, x_n$ are given set of observations which are need not be equally spaced and let $y_0, y_1, y_2, \dots, y_n$ are their corresponding values, where $y = f(x)$ be the given function

then
$$f(x) = \frac{(x-x_1)(x-x_2)\dots(x-x_n)}{(x_0-x_1)(x_0-x_2)\dots(x_0-x_n)} y_0 + \frac{(x-x_0)(x-x_2)\dots(x-x_n)}{(x_1-x_0)(x_1-x_2)\dots(x_1-x_n)} y_1 + \dots + \frac{(x-x_0)(x-x_1)\dots(x-x_{n-1})}{(x_n-x_0)(x_n-x_1)\dots(x_n-x_{n-1})} y_n$$