UNIT- IV

Probability Distributions and Graphs

Normal Probability Curve. Meaning of probability- Principles of normal curve – Properties of normal curve. Divergence from normality – Skewness and Kurtosis. Grading Scales. Graphical Representation in Statistics; Line diagram, Bar diagram, Histogram, Frequency Polygon, Ogive Curve and Pie chart.

Probability Examples

- A jar contains 30 red marbles, 12 yellow marbles, 8 green marbles and 5 blue marbles

---

- What is the probability that you draw and replace marbles 3 times and you get NO red marbles?

  - There are 55 marbles, 25 of which are not red
  - P(getting a color other than red) = P(25/55) = .455
  - Probability of this happening 3 times in a row is found by .455*.455*.455 = .094

PRINCIPLES AND PROPERTIES OF NORMAL CURVE

- Symmetry
- Bell-Shaped Curve
- Mean, Median, and Mode Equality
- Standard Deviation
- Z-Scores

SKEWNESS

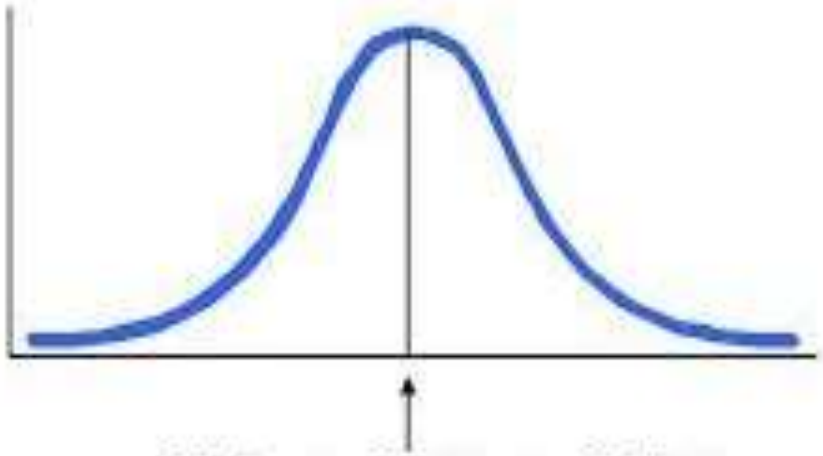
- Skewness is a statistical measure that describes the asymmetry of a probability distribution or a set of data points.
- In simpler terms, it quantifies the extent and direction of skew (departure from horizontal symmetry) in a dataset.

- **Positive Skewness:**

- If the distribution of data is skewed to the right, it has a long right tail.
- This means that the majority of the data points are concentrated on the left side, and there are a few extremely high values on the right side.

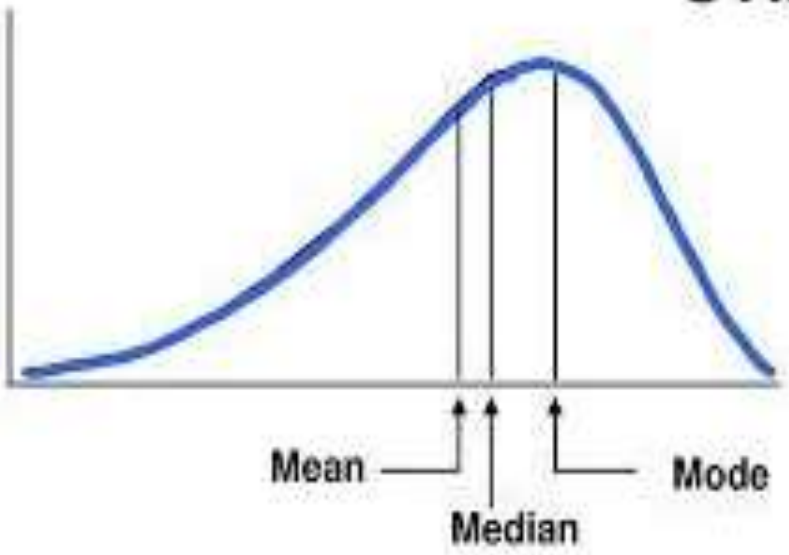
- **Negative Skewness:**

- If the distribution of data is skewed to the left, it has a long left tail.
- This indicates that the majority of data points are concentrated on the right side, and there are a few extremely low values on the left side.

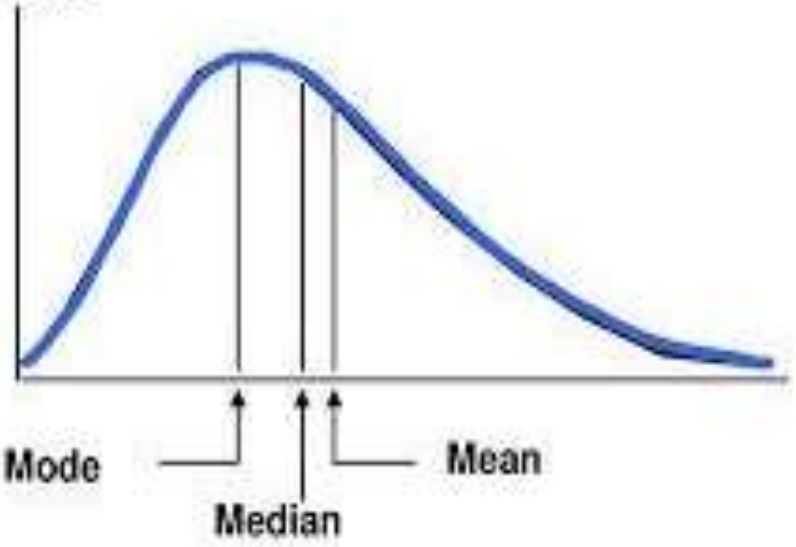


Mode = Mean = Median

SYMMETRIC



SKEWED LEFT
(negatively)

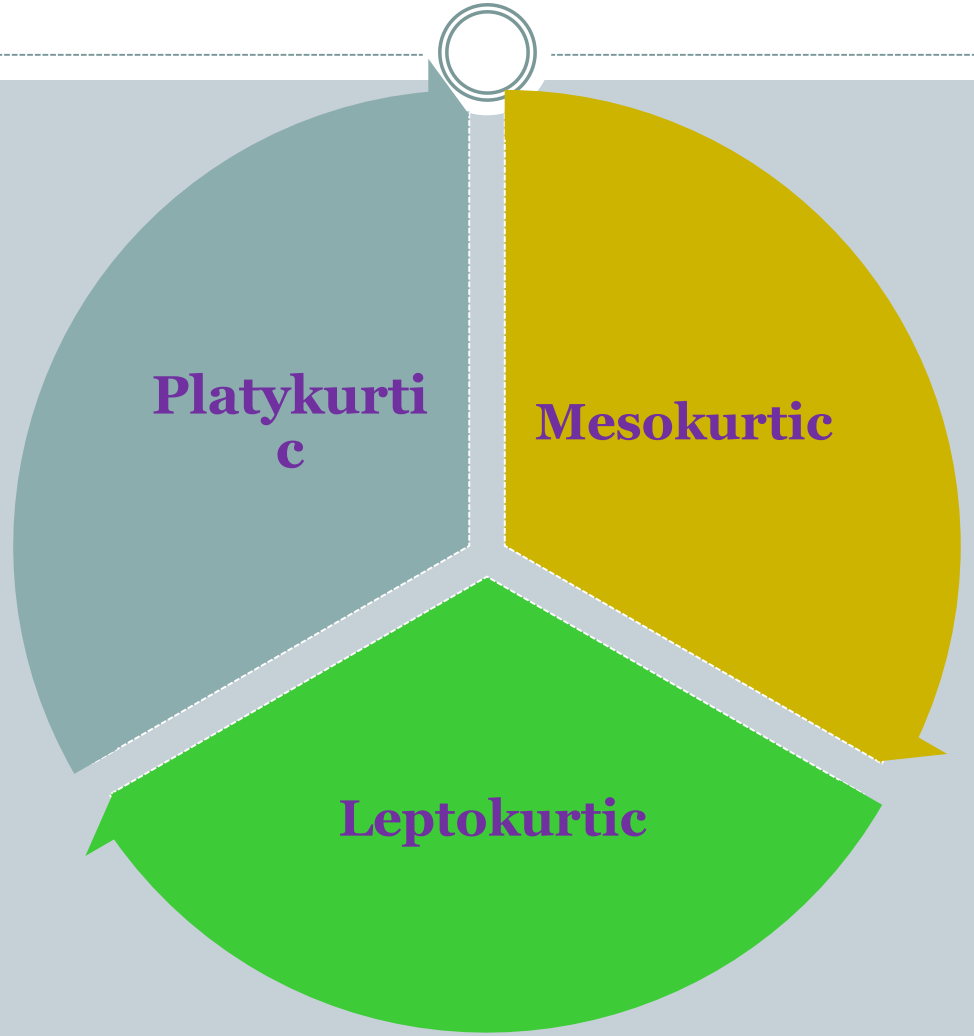


SKEWED RIGHT
(positively)

KURTOSIS

- Kurtosis is a statistical measure that describes the shape, or peakedness, of a probability distribution or a set of data points.
- It provides information about the tails of a distribution and how much data is concentrated near the mean.
- In particular, kurtosis helps to identify whether the tails of a distribution are heavy or light compared to a normal distribution.

Types of kurtosis

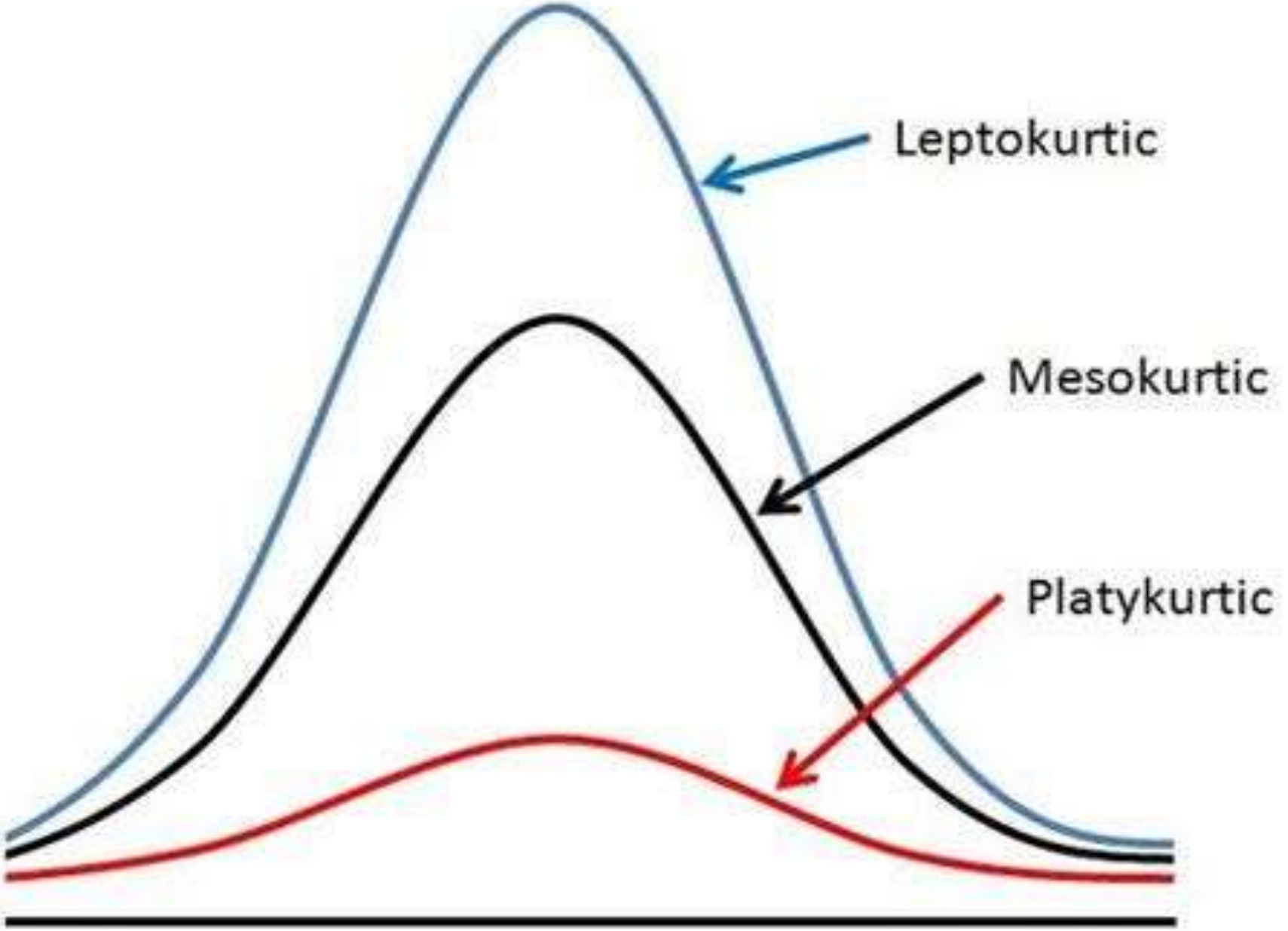


- **Mesokurtic (Normal Kurtosis):**

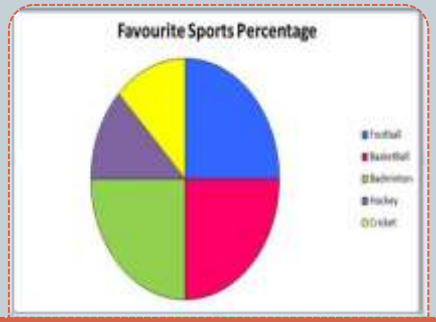
- A mesokurtic distribution has kurtosis equal to 0.
- This means that its tails are similar to those of a normal distribution.
- The majority of statistical distributions, including the normal distribution, fall into this category.

- **Leptokurtic (Excess Kurtosis > 0):**
- A leptokurtic distribution has "fat" tails, indicating that it has more extreme values than a normal distribution.
- The excess kurtosis value (kurtosis minus 3) is positive.
- Leptokurtic distributions are more peaked at the center and have heavier tails.

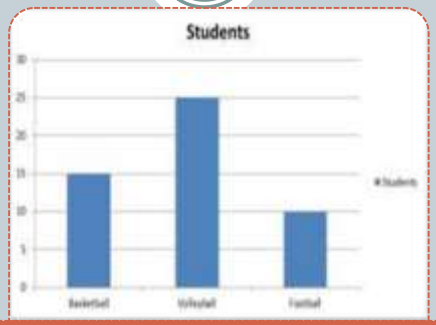
- **Platykurtic (Excess Kurtosis < 0):**
- A platykurtic distribution has "thin" tails, indicating that it has fewer extreme values than a normal distribution.
- The excess kurtosis value is negative.
- Platykurtic distributions are less peaked at the center and have lighter tails.


Leptokurtic
Mesokurtic
Platykurtic

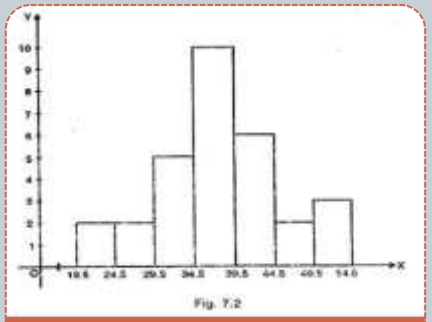
Graphical Representation in Statistics



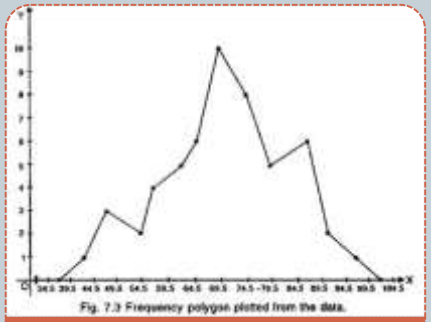
Pie Graph



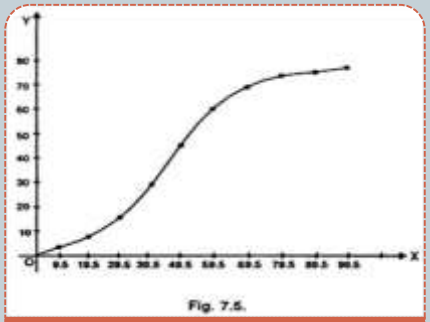
Bar Graph



Histogram



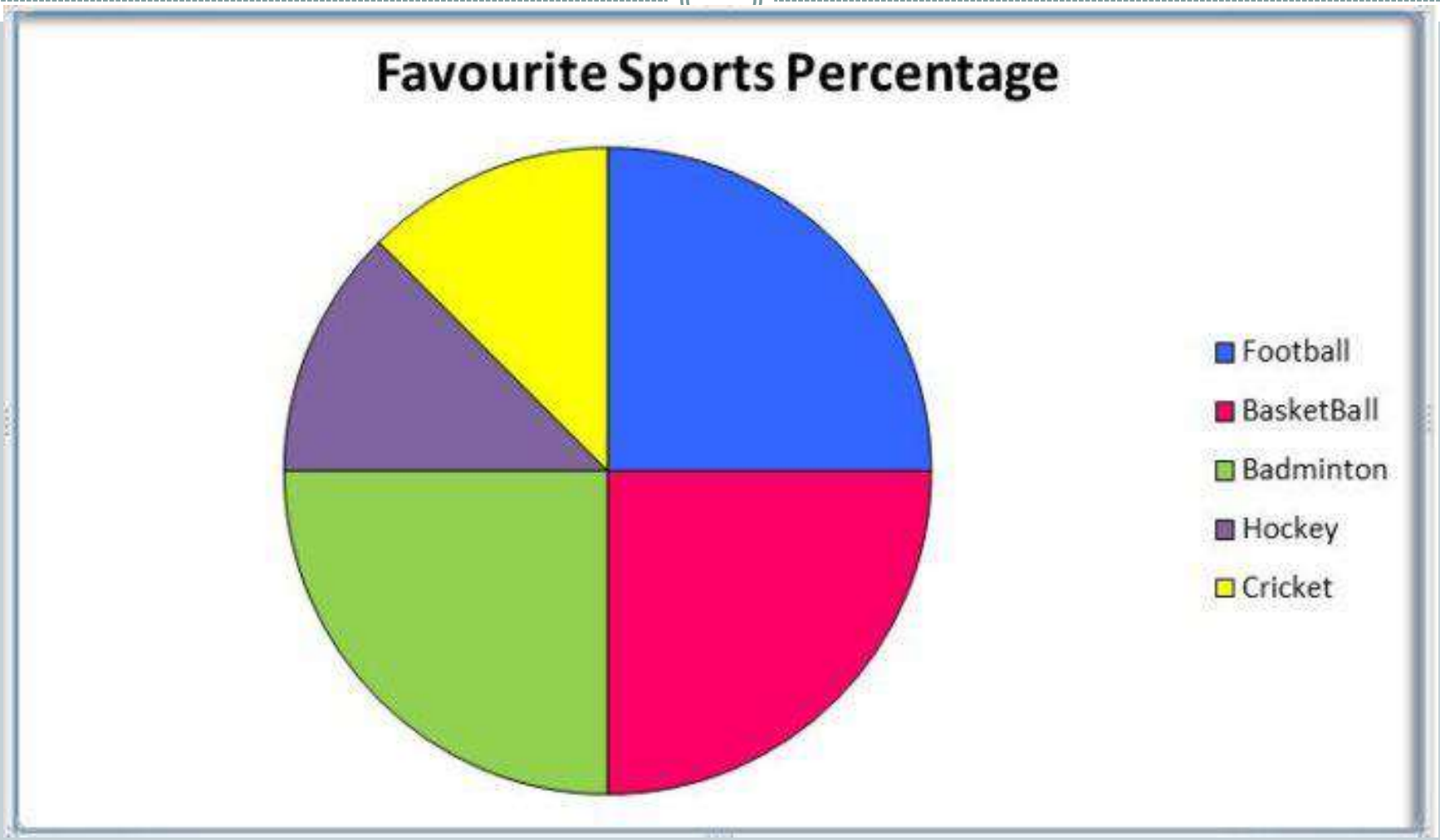
Frequency Polygon



Ogive Curve

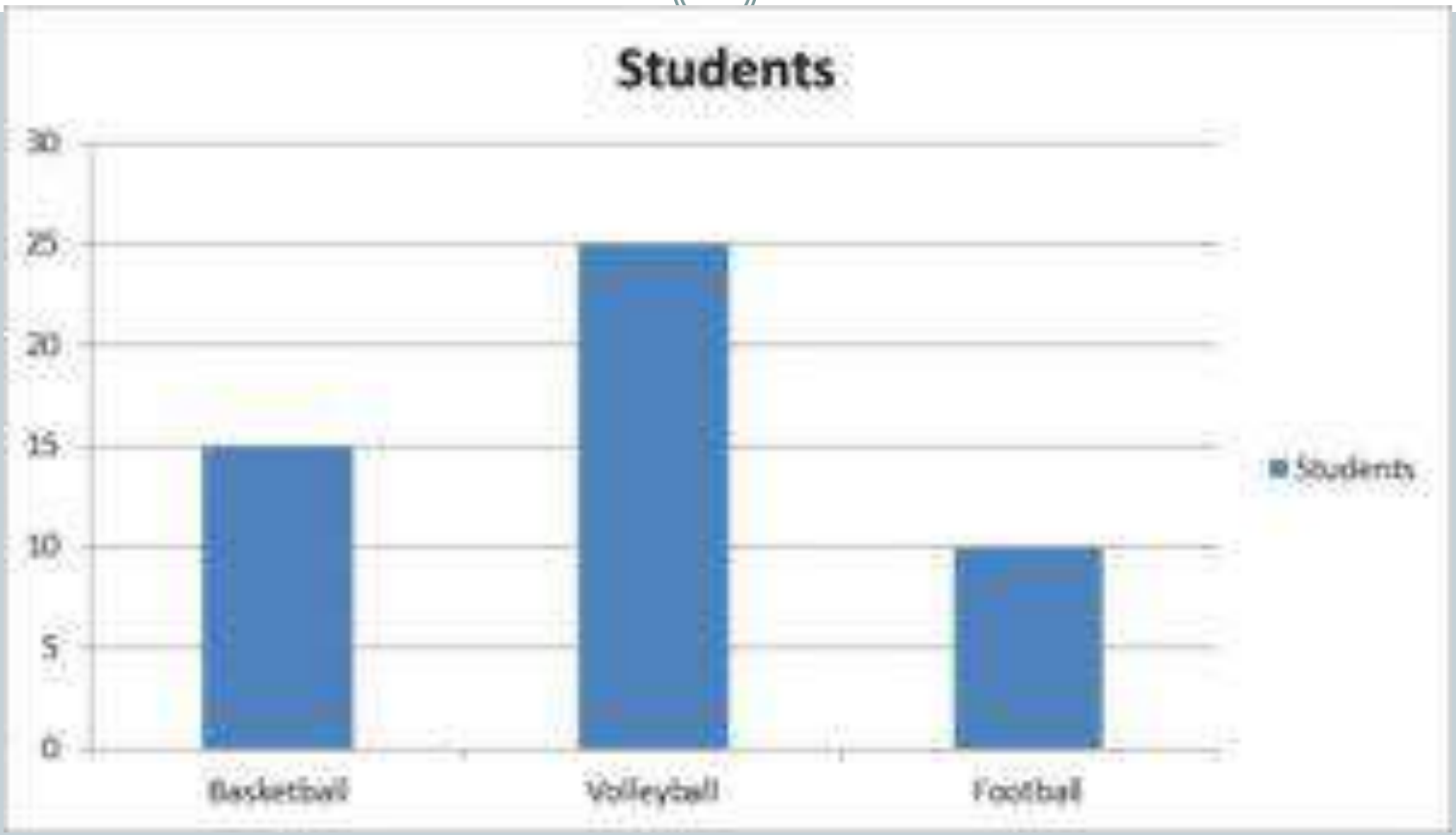
Pie Graph

- 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.


Favourite Sports Percentage
Football
BasketBall
Badminton
Hockey
Cricket

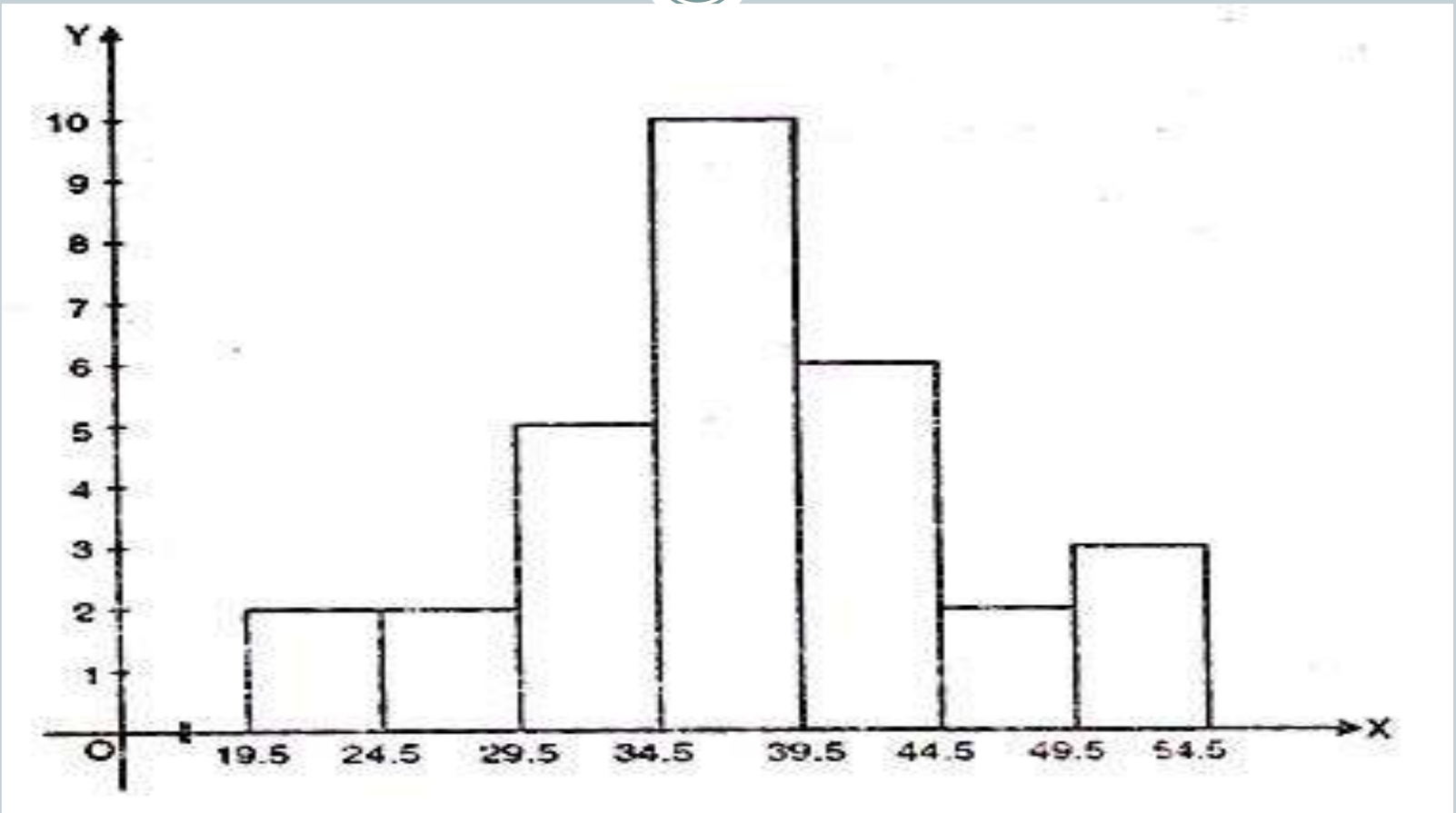
Bar Diagram

- 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.


Students
30
25
20
15
10
5
0
Basketball
Volleyball
Football
Students

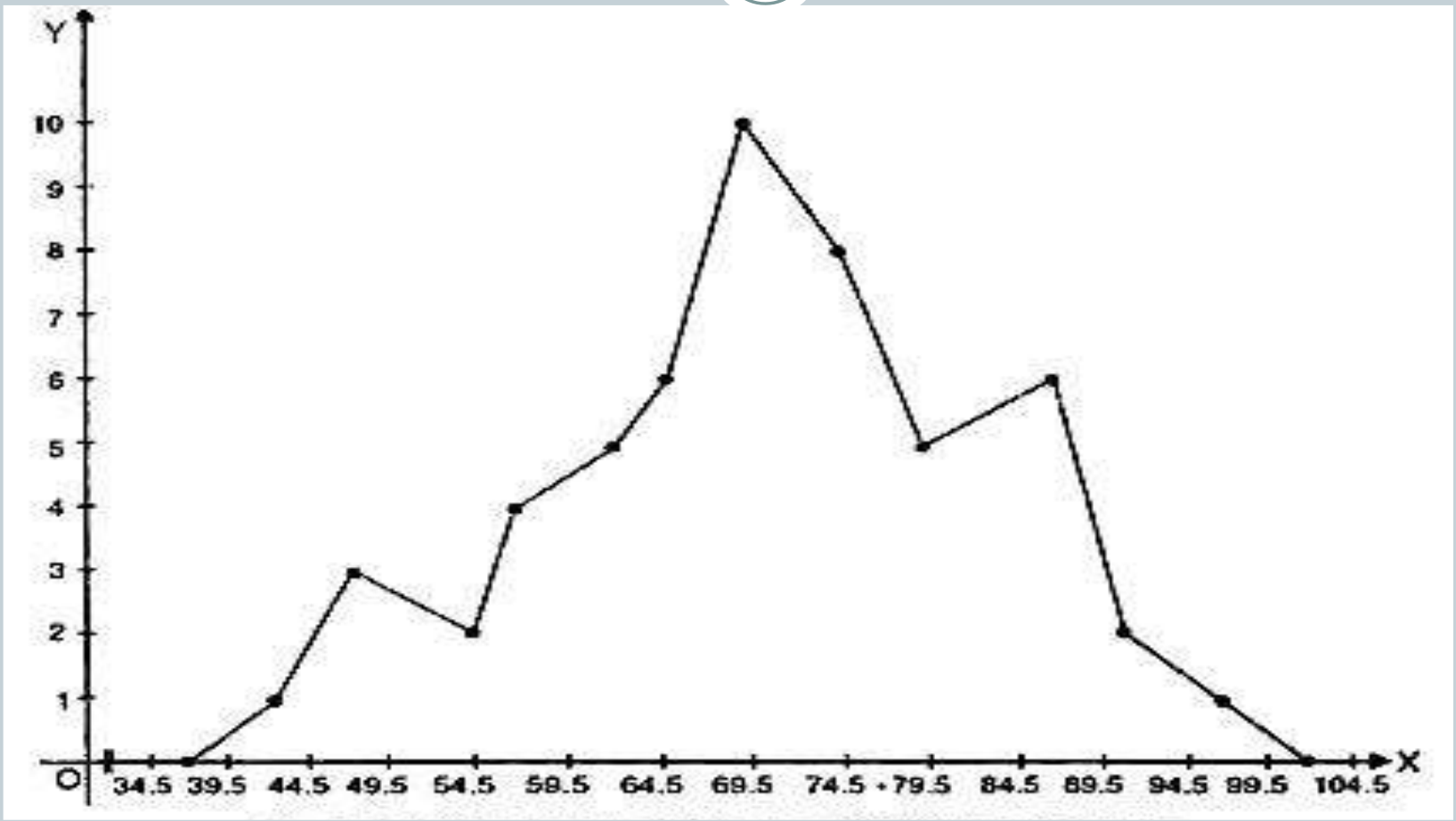
Histogram

- 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.

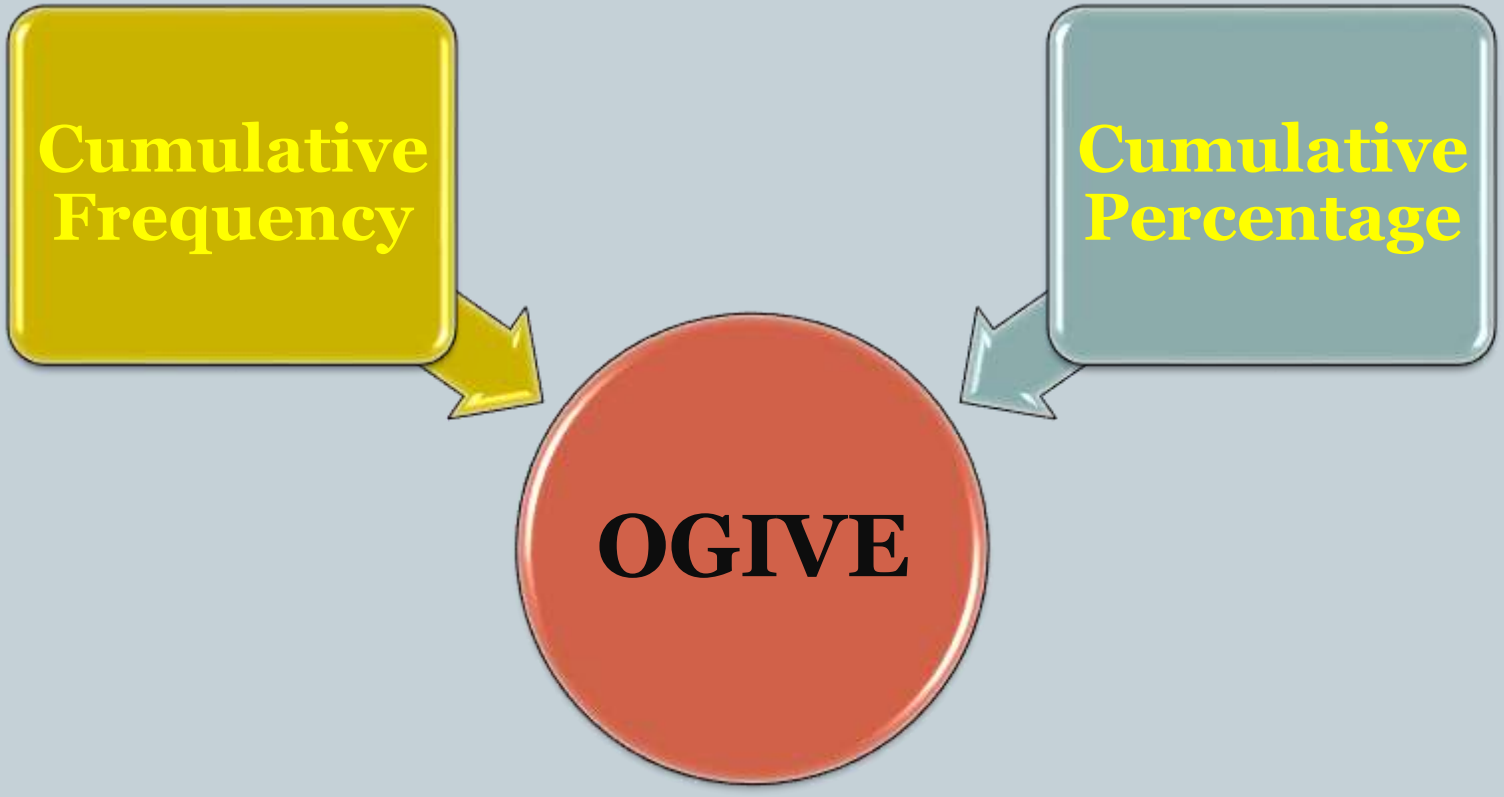

Y
10
9
8
7
6
5
4
3
2
1
O
19.5
24.5
29.5
34.5
39.5
44.5
49.5
54.5
X

Frequency Polygon

- A frequency polygon is a graphical representation of a **frequency distribution**.
- It is created by connecting the midpoints of the intervals in a histogram with straight line segments.
- This type of graph is particularly useful for displaying the shape and central tendency of a dataset.

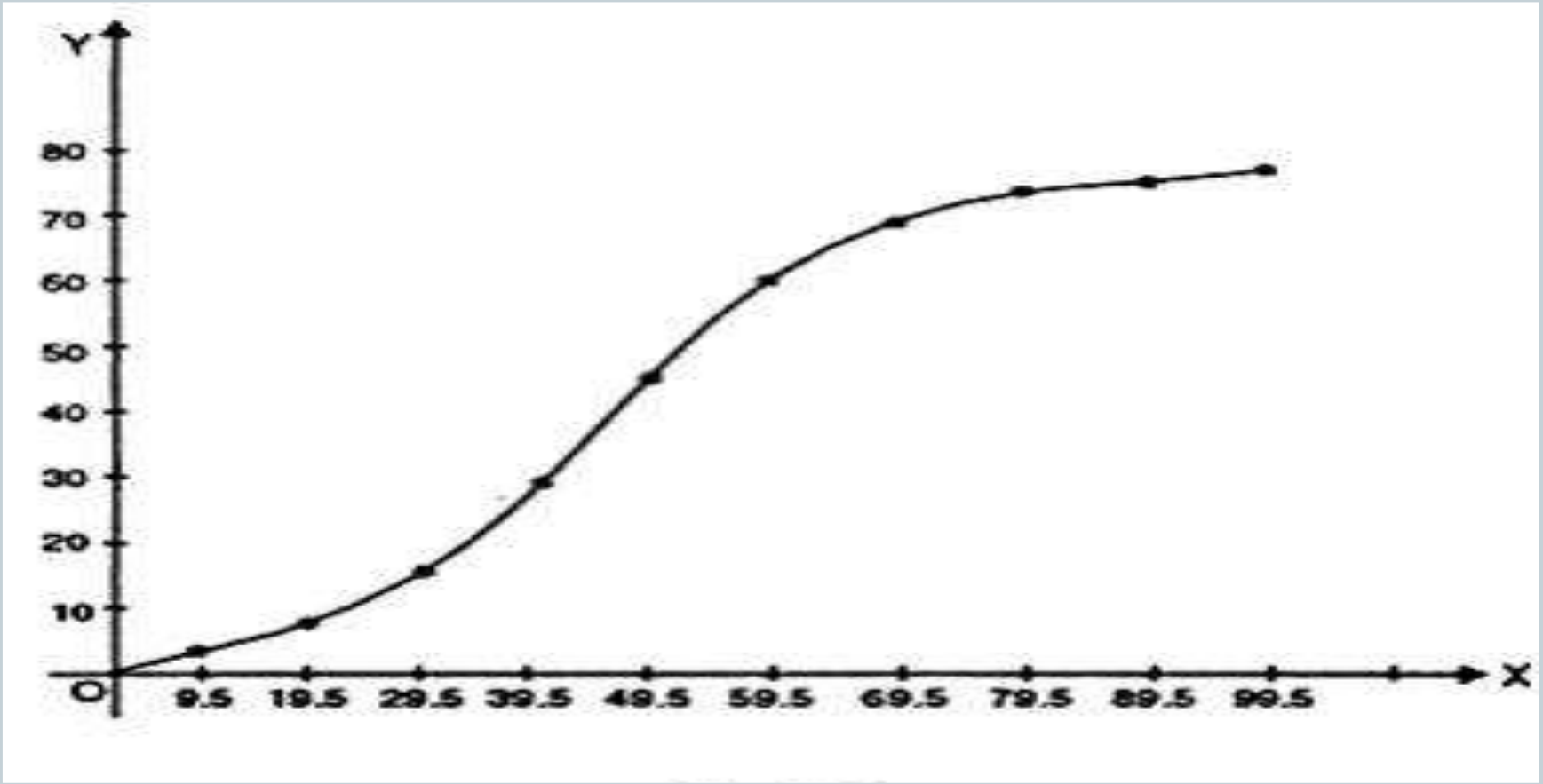

Y
10
9
8
7
6
5
4
3
2
1
O
34.5
39.5
44.5
49.5
54.5
59.5
64.5
69.5
74.5
79.5
84.5
89.5
94.5
99.5
104.5
X

Ogive



Cumulative Frequency Curve

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


Y
80
70
60
50
40
30
20
10
O
9.5
19.5
29.5
39.5
49.5
59.5
69.5
79.5
89.5
99.5
X

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.

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