Histograms and Polygons

EGI SAFITRI, S.MAT., M.SI

FREQUENCY DISTRIBUTION

Example:There are 80 calves weight data as listed below (in kg)

79	92	49	93	48	76	74	71	81	90	98	72	87	67	80	75
80	80	84	91	90	61	70	72	91	97	93	91	82	88	70	81
70	70	71	74	92	99	38	95	56	80	81	59	74	71	73	77
68	63	72	60	85	83	51	82	65	60	93	67	83	89	86	63
90	76	35	63	83	88	73	70	74	66	43	88	86	79	68	75

Step-by-step creation

Determine the range, which is the largest data minus the smallest data: 99 - 35 = 64
Determine the number of interval classes required. Generally between 5 - 15. Sturges method/rule:
Number of classes: 1 + (3.3) log n where n = 80
= 1 + 3.3 x log 80
= 1 + 3.3 x 1.9031
= 7.2802
Many classes can be 7 classes

Continued

The class length (p) is determined by the rule:

P = range / number of classes

= 64/7 = 9.14

so that it can be taken number 9.

Select the lower end of the first interval class. For that, it can be taken equal to the smallest data or data values that are smaller than the smallest data but the difference must be less than the predetermined class length.

Create a frequencies distribution list

weight(kg)	frequency (f)
31 - 40	2
41 - 50	3
51 - 60	5
61 - 70	14
71 - 80	24
81 - 90	20
91 - 100	12

weight(kg)	frequency (f)
35 - 44	3
45 - 54	3
55 - 64	8
65 - 74	23
75 - 84	20
85 - 94	19
95 - 104	4

Relative frequency distribution

If the absolute frequency distribution is expressed in percent

weight(kg)	Absolut frequency	f relatif (%)
31 - 40	2	2.5
41 - 50	3	3.75
51 - 60	5	6.25
61 - 70	14	17.5
71 - 80	24	30
81 - 90	20	25
91 - 100	12	15
Jumlah	80	100

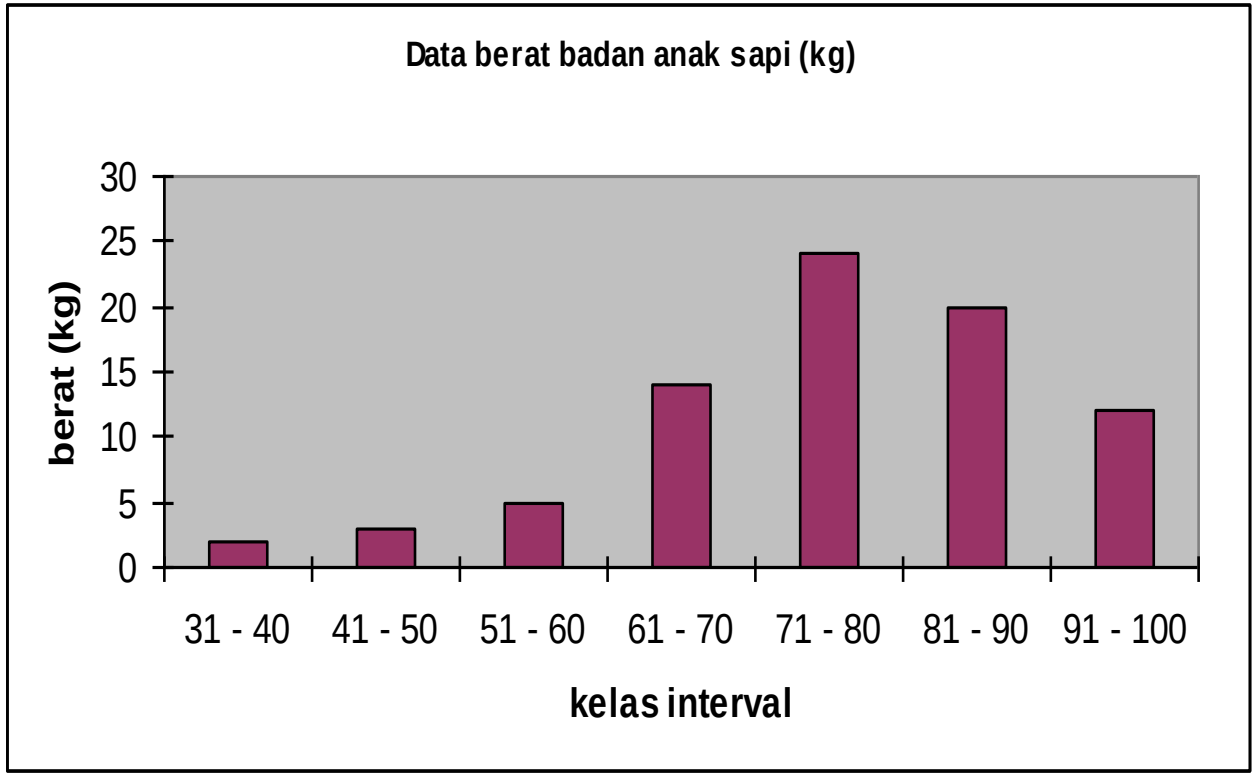
HISTOGRAM AND FREQUENCY POLYGON

Histogram: a diagram of a frequency distribution

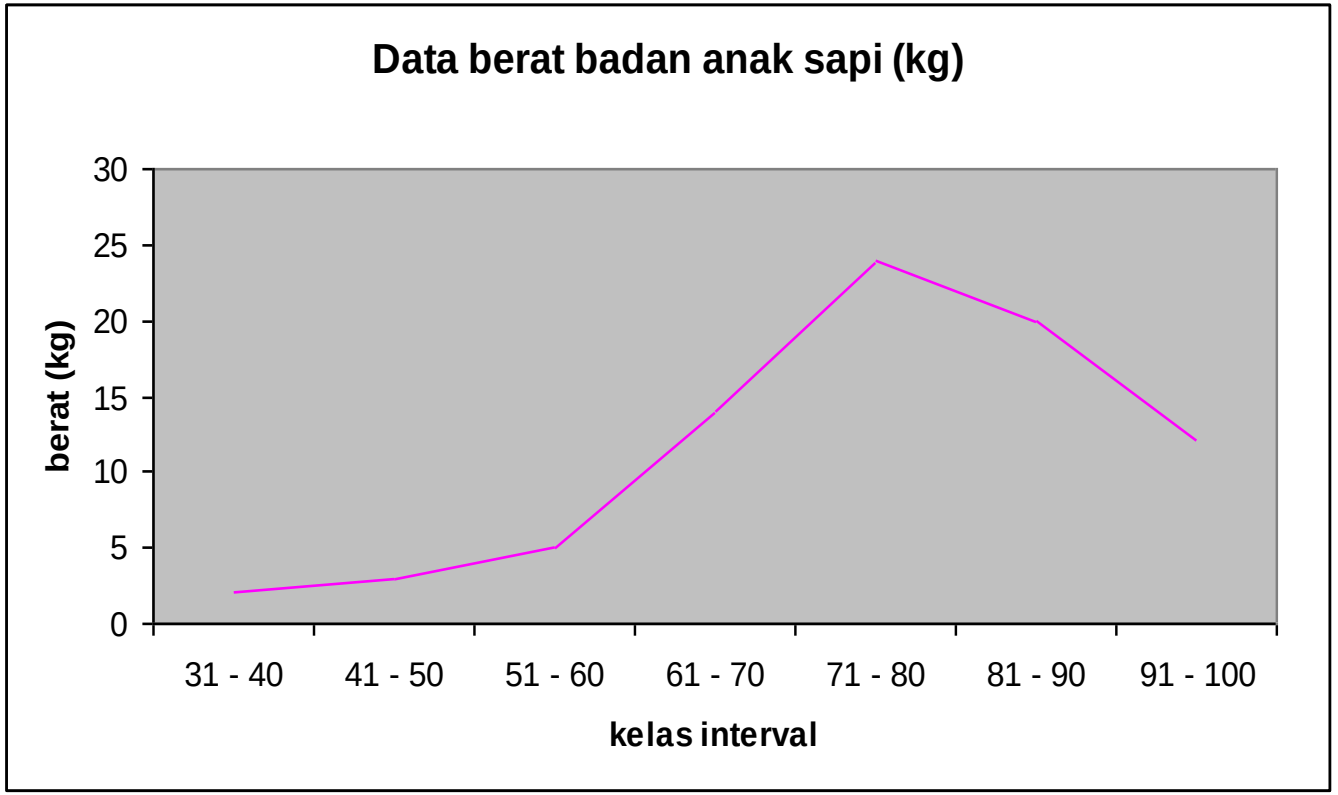
Frequency

Polygon: A line connecting half the interval distances on the flat axis of a histogram.

HISTOGRAMS

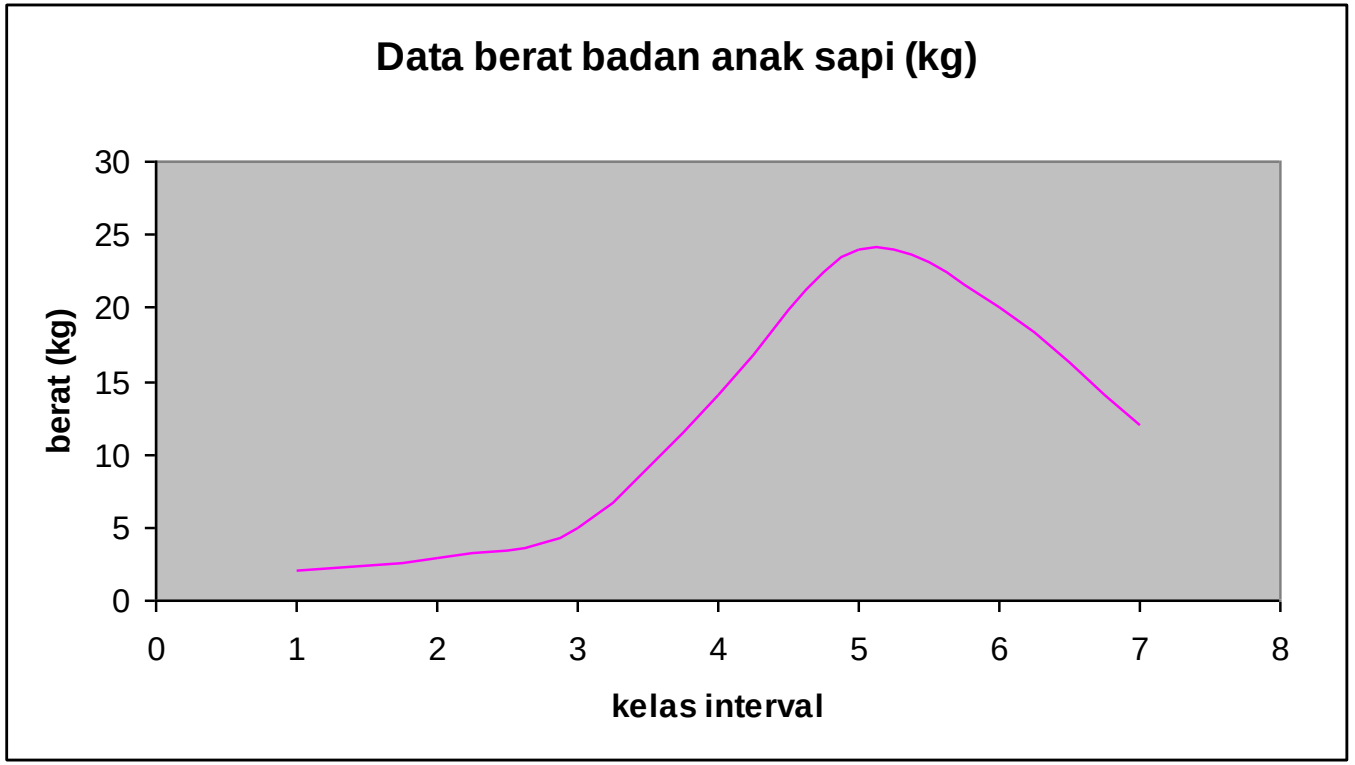

Data berat badan anak sapi (kg)
berat (kg)
30
25
20
15
10
5
0
31 - 40
41 - 50
51 - 60
61 - 70
71 - 80
81 - 90
91 - 100
kelas interval

FREQUENCY POLYGON



FREQUENCY CURVE

Poligon frequensi yang merupakan garis patah-patah dapat didekati oleh sebuah lengkungan halus yang bentuknya secocok mungkin dengan poligon frequensi. Lengkungan tersebut dinamakan kurva frequensi



POPULATION MODEL

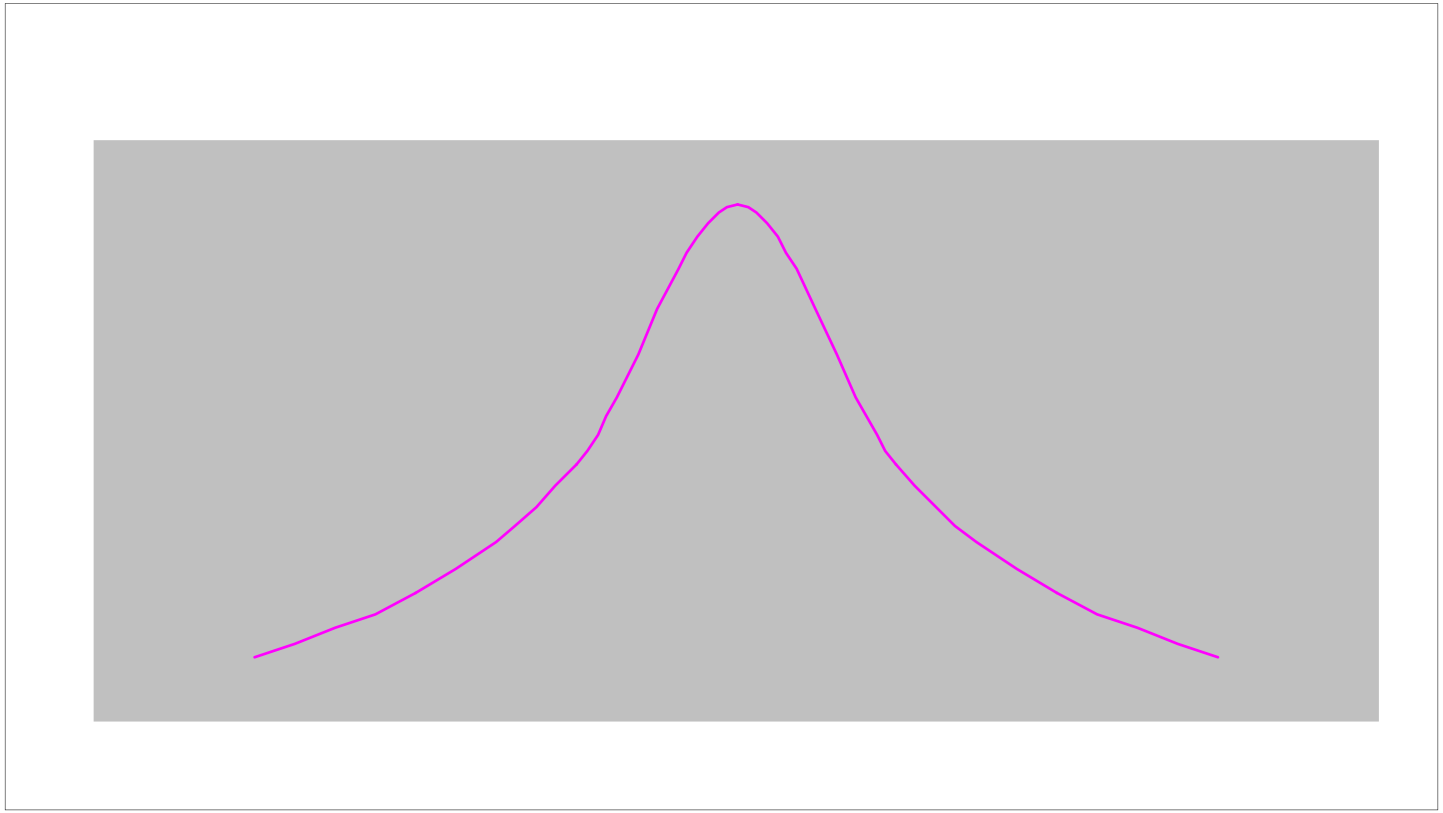
The frequency curve obtained from a representative sample taken from the population is called the population model.

EXAMPLES

- Normal model
- Symmetric model
- Positive or left-skewed model
- Negative or right-skewed model
- J-shape model
- U-shape model

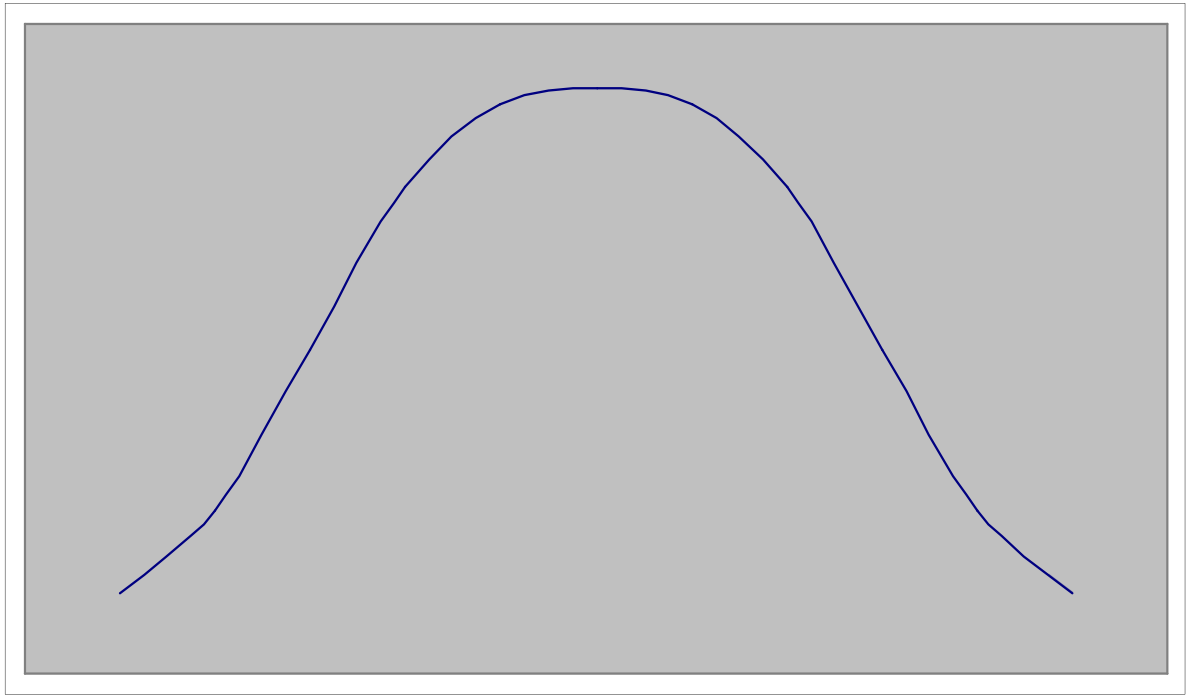
NORMAL MODEL

The normal model is more accurately described by its mathematical equation. The shape of the normal model is always symmetric and has a peak. A curve with a peak is called unimodal



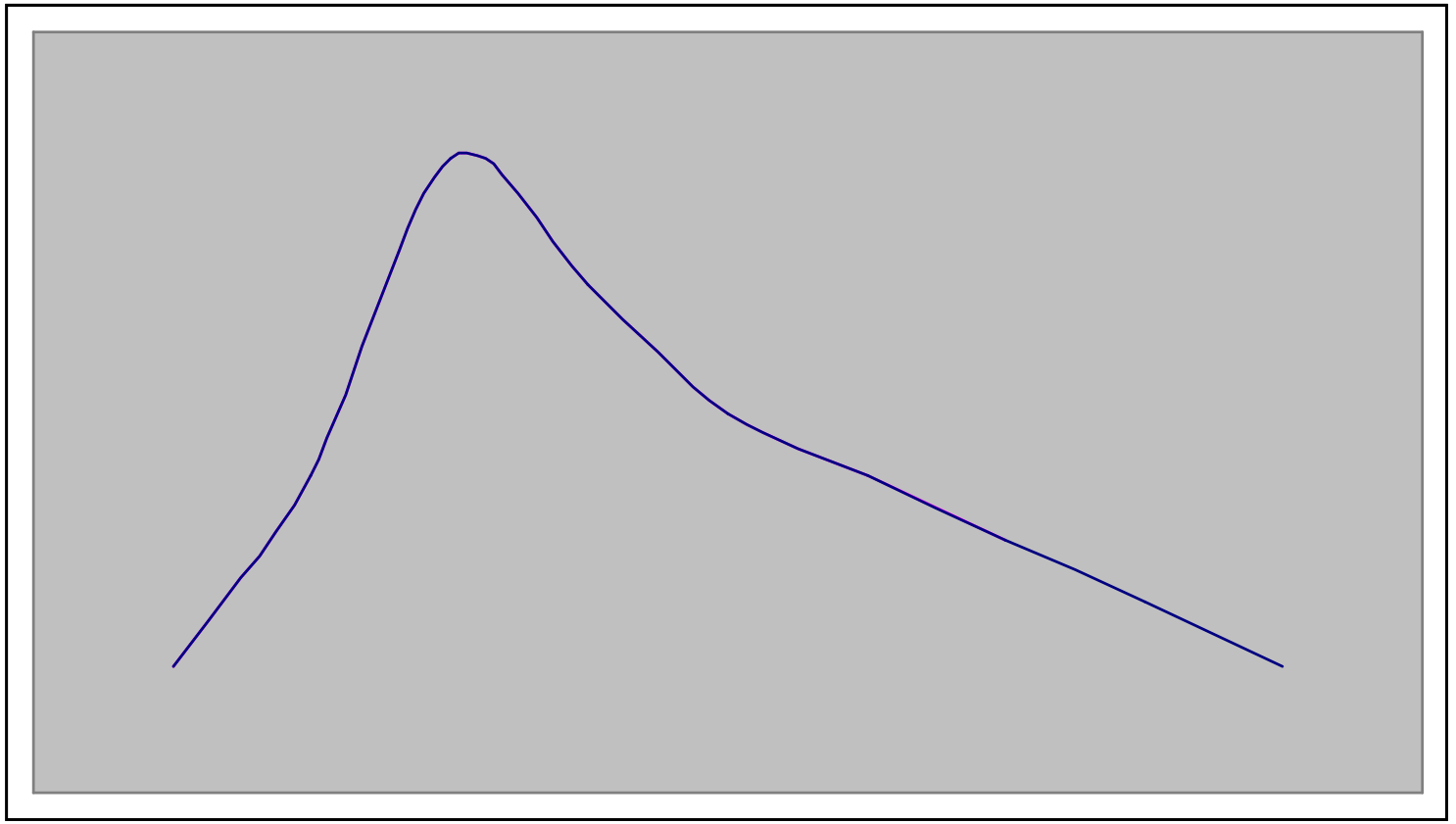
SIMETRIK MODEL

The normal model is always symmetric but not vice versa, here the peak is also unimodal



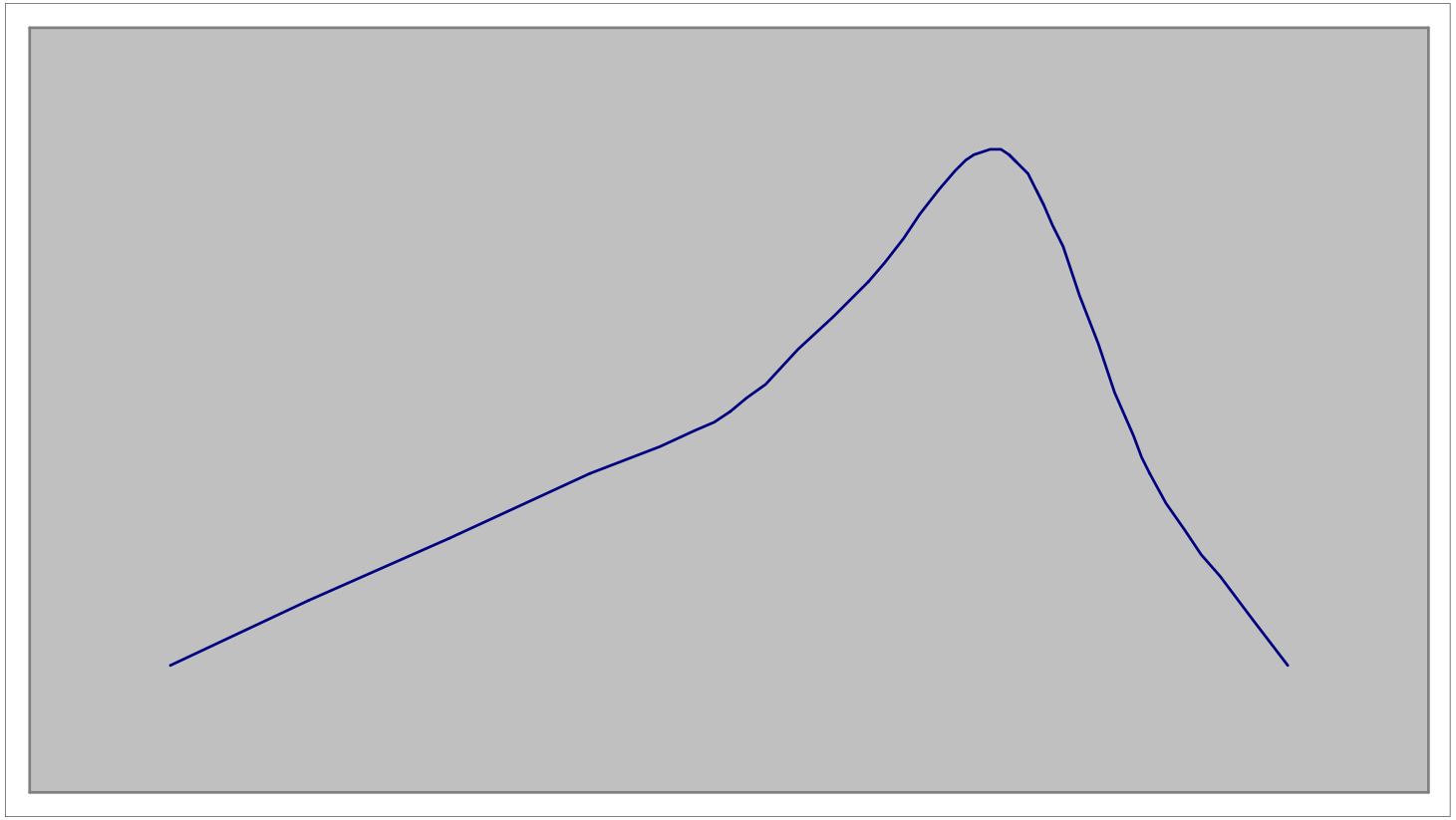
POSITIVE MODEL

The positive model describes that there are few symptoms with increasing values



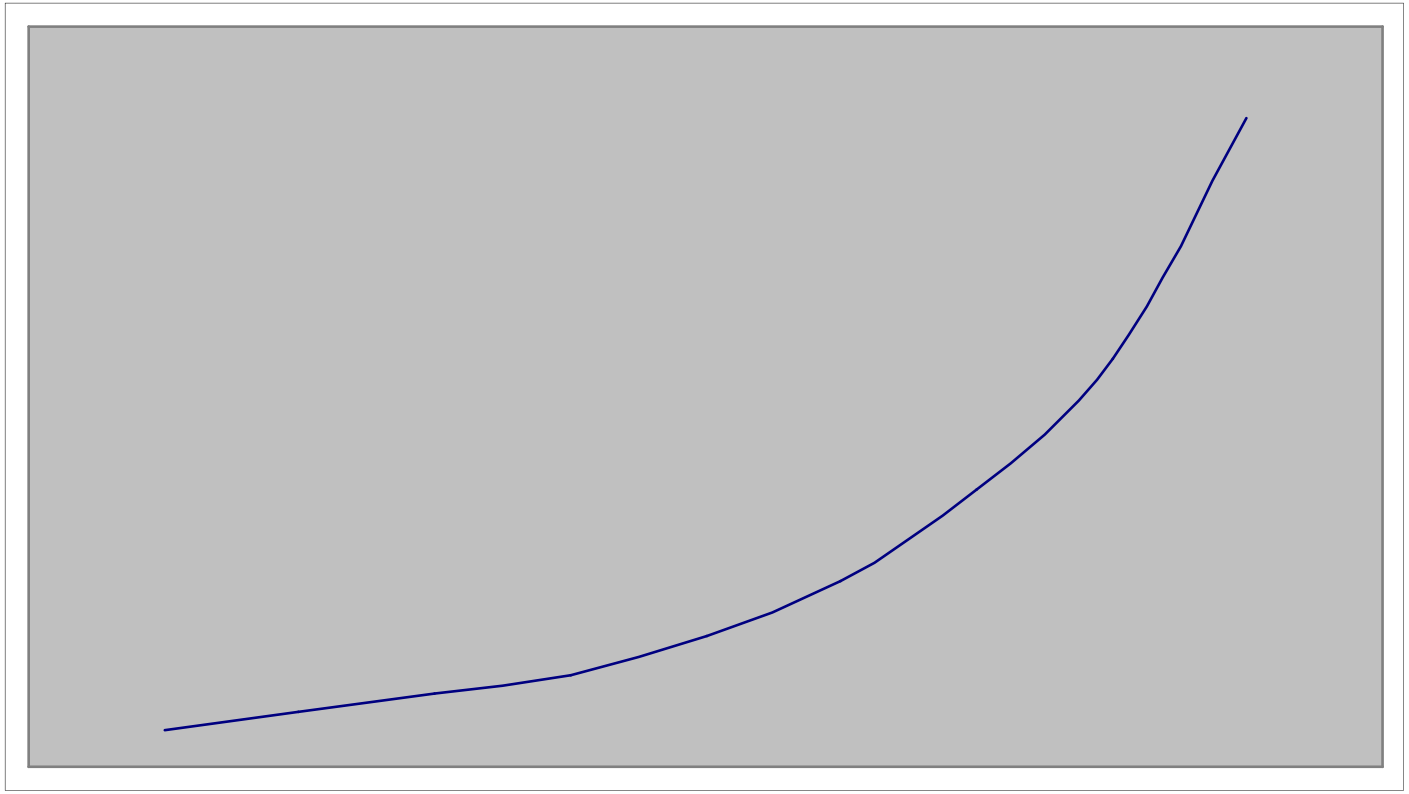
NEGATIVE MODEL

The negative model describes that there are many symptoms with increasing values



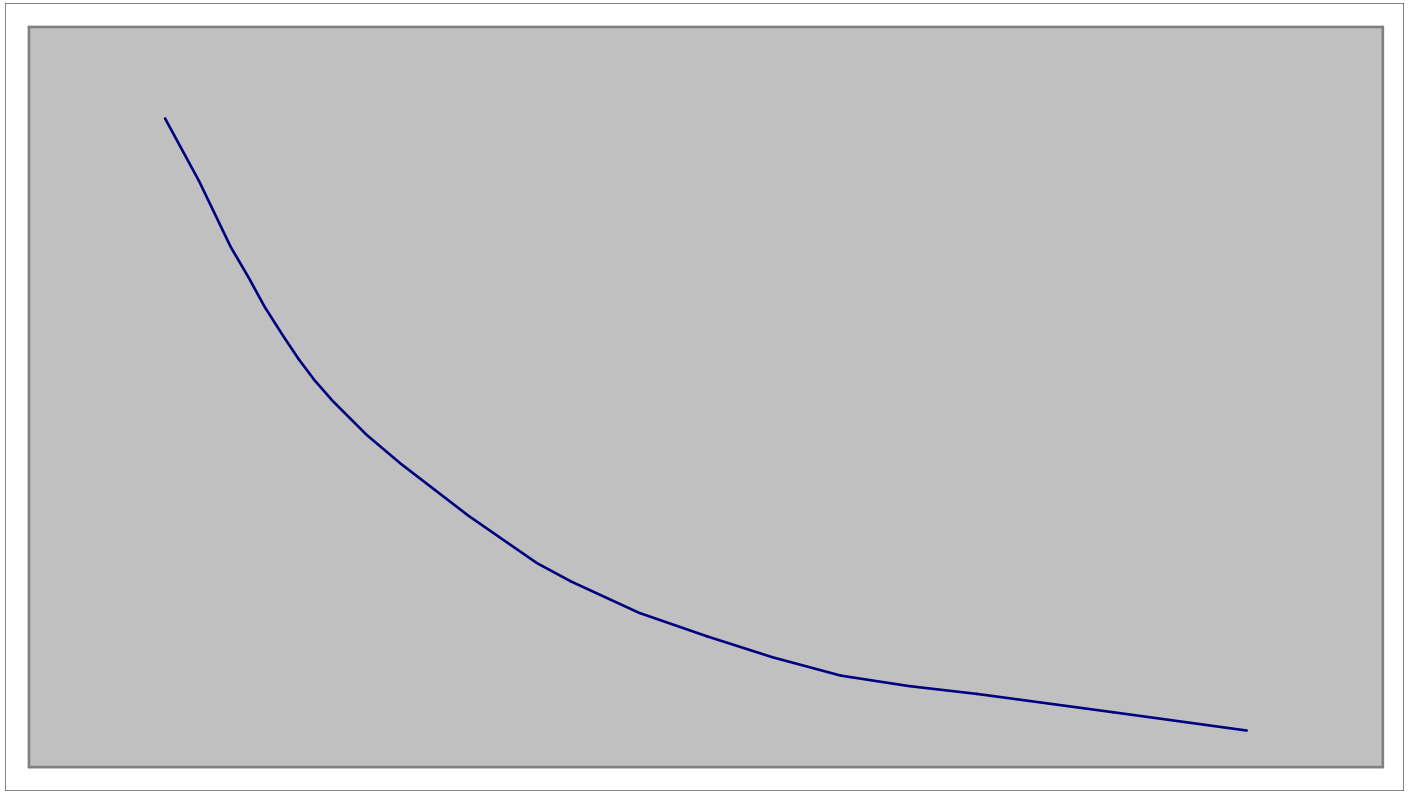
J MODEL (Exponential)

The J model is found in economics, industry, livestock and fisheries



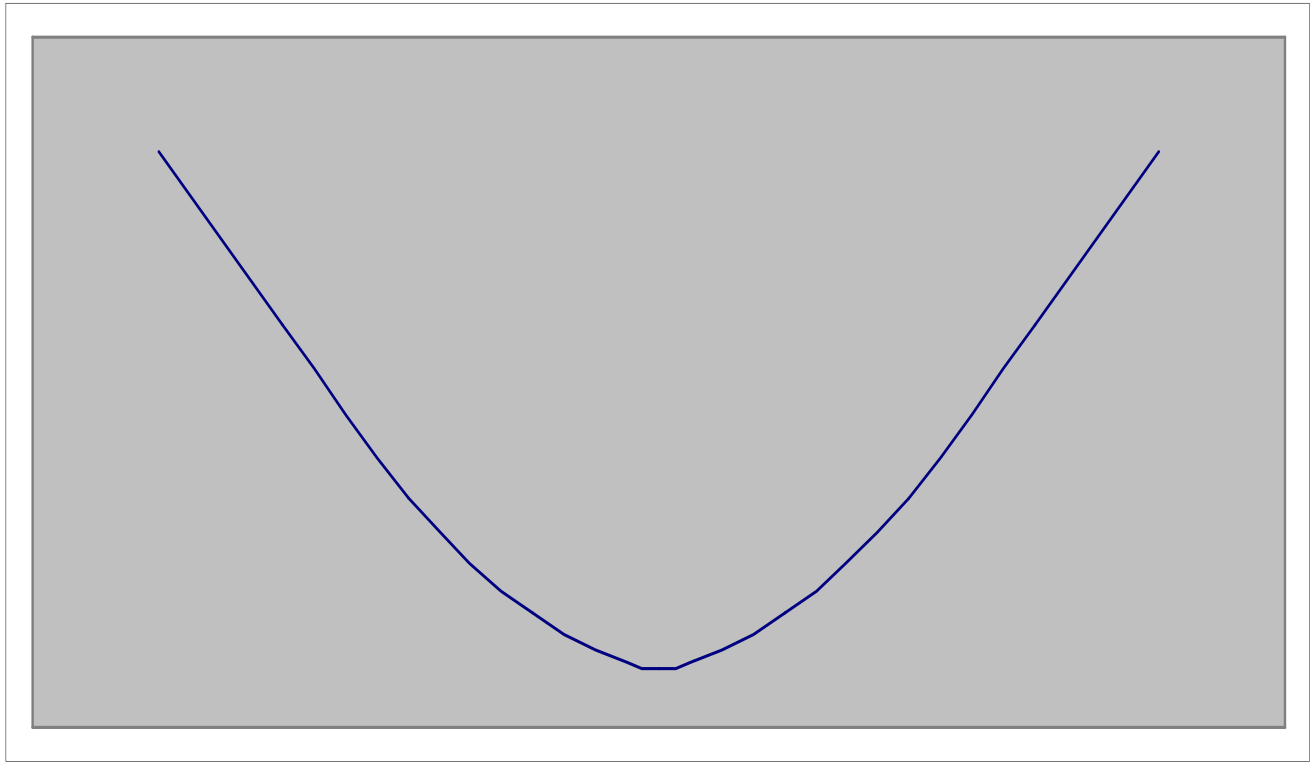
Reversed J MODEL

The J model is found in economics, industry, livestock and fisheries



U MODEL

The U model has two peaks so it is called bimodal. Models with more than one peak are called multimodal



Frequency Polygon

A frequency polygon is a statistical tool equivalent to a Histogram used to represent and compare data when given in the form of cumulative frequencies. Frequency polygons use line graphs to represent quantitative data. Frequency polygons are one of the good methods to represent statistical data for easy reading. In statistics, we handle a lot of data, and reading it quickly is necessary to solve statistical problems effectively.

Frequency Polygons in Statistics

A visual representation of the frequency distribution of continuous data is the frequency polygon. Karl Pearson, a British statistician, first described it in the late 19th century. Based on the research of previous statisticians such as Francis Galton and Adolphe Quetelet, he came up with the idea. Their value comes from their ability to graphically depict data, making it easier to understand and analyze. Therefore, they serve as an important tool in statistics, assisting researchers in finding patterns and trends in very large data sets.

They are useful for displaying data sets with many observations or values, which is one of their main features. Polygons represent the distribution of data concisely and clearly by dividing the data into equal intervals and charting the frequency of each container. Since they are points that fall outside the range of a typical distribution, they are also useful in finding outliers or abnormalities in the data set. Again, this makes it easy to identify and examine odd or unexpected data points.

Definition of Frequency Polygon

A frequency polygon is a type of graphical data distribution that helps recognize data by giving it a specific shape. Although frequency polygons and histograms are very similar, they are more effective when comparing two or more sets of data. As a line graph, it mainly displays data from a cumulative frequency distribution.

Formula for Finding the Midpoint of the Frequency Polygon

If you want to graph a frequency polygon, you need to know the midpoint or class mark for each class interval. Here's the formula to achieve that:

$$\text{Class Value (Midpoint)} = (\text{Upper Bound} + \text{Lower Bound}) / 2$$

Frequency Polygon Graph

A Frequency Polygon Graph is a graphical representation of data given in the form of class intervals and frequencies. Let us consider an example for better understanding.

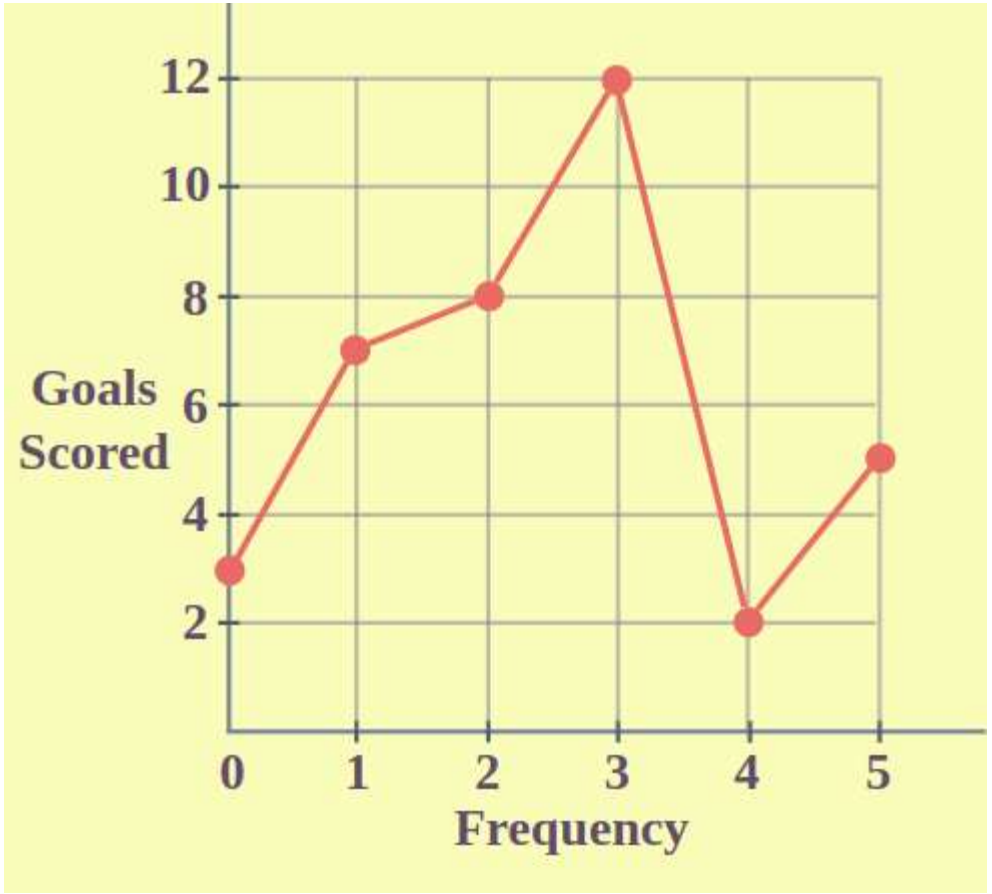
Example

Draw a Frequency Polygon graph for the following data representing the number of goals scored in a match in a league throughout the season

Goals Scored	Frequency
0	3
1	7
2	8
3	12
4	2
5	5

Example

For a given data, we can plot a frequency polygon by representing the scored goal on the vertical axis and the frequency on the horizontal axis, as follows:



Cumulative Frequency Polygon

In statistics, the cumulative frequency of a data set is displayed graphically using a cumulative frequency polygon. The points at the upper class boundary are plotted against the corresponding cumulative frequency to create a graph. The graph illustrates the visual accumulation of data over time or intervals. The polygons created by lines connecting the points help visualize trends and patterns in the data. It helps analysts understand the structure and main tendencies of cumulative frequency distributions by showing them clearly. Researchers and analysts can gain important insights into the distribution of data and make wise judgments based on the trends seen by using cumulative frequency polygons.

How to Draw Frequency Polygons?

Following the procedure below, make a histogram before starting to draw the frequency polygon:

Step 1: First, choose the class interval and then indicate the value on the axis.

Step 2: Label the horizontal axis with the midpoint of each interval.

Step 3: Label the vertical axis with the frequency of the class.

Step 4: Mark a point at the height in the middle of each class interval corresponding to the frequency of each interval.

Step 5: Use a line segment to connect these points.

Step 6: The representation obtained is a frequency polygon.

Histogram and Frequency Polygon

The differences between frequency polygons and histograms are discussed in the table below:

Frequency Polygon	Histogram
The frequency polygon graph allows for more accurate visual data comparisons.	In a histogram graph, data comparison is not visually pleasant.
The class midpoint is used in the frequency polygon graph.	Frequency in the histogram is distributed evenly across the class intervals.
Data from a specific class interval is represented by the correct points in the frequency polygon graph.	The height of the bars in the histogram only represents the amount of data.
The line segments used to express a curve are called the frequency polygon graph.	A histogram is a graph that displays data as a series of continuous vertical bars.

Frequency Polygon Example Problem

Example 1: The frequency data beside this is used to create frequency polygons.

Class Interval	Frequency
49.5-59.5	5
59.5-69.5	10
69.5-79.5	30
79.5-89.5	40
89.5-99.5	15

By first determining the classmark using the equation

Classmark = (Upper Limit + Lower Limit)/2,

we can construct a frequency polygon without a histogram. Also, by combining the previous and next frequencies, we can obtain the cumulative frequency of each class interval.

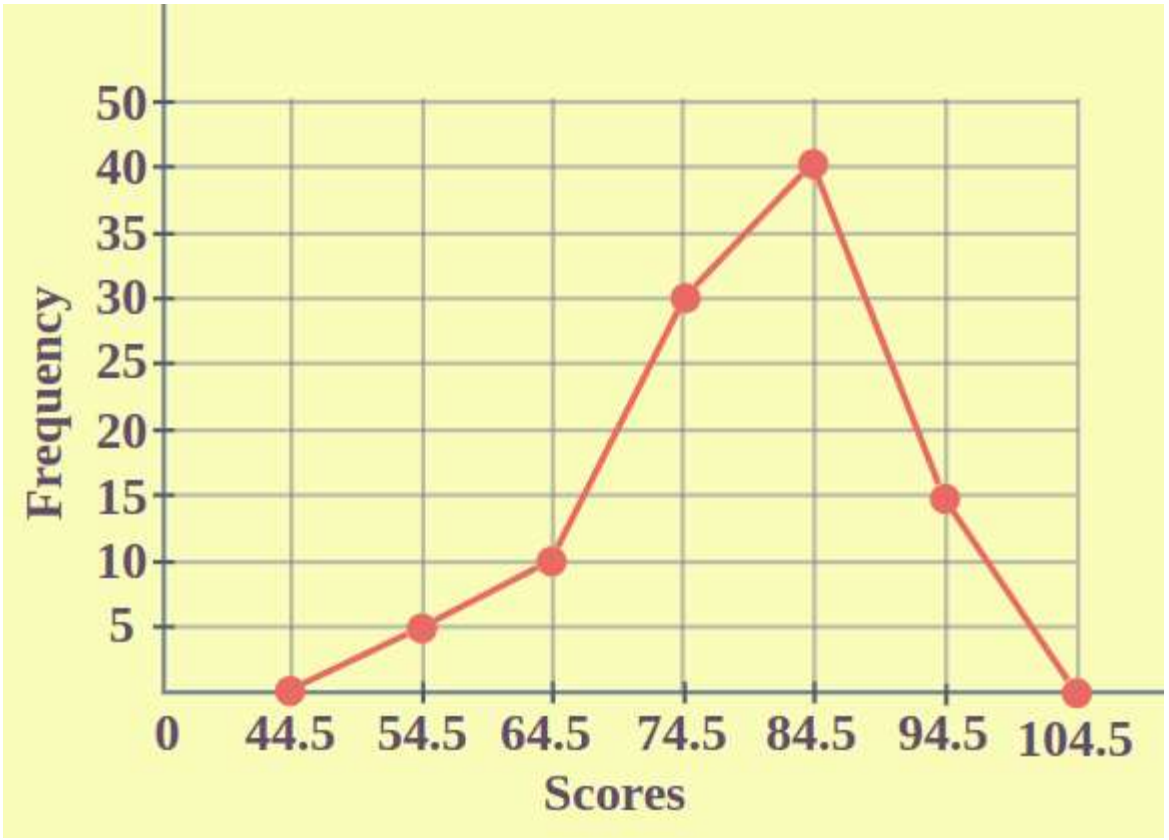
Class Interval:

- (59.5 + 49.5/2) = 54.5
- (69.5 + 59.5/2) = 64.5
- (79.5 + 69.5/2) = 74.5
- (89.5 + 79.5/2) = 84.5
- (99.5 + 89.5/2) = 94.5

Class Interval	Lower Bound	Upper Bound	Class Mark	Frequency
49.5-59.5	49.5	59.5	54.5	5
59.5-69.5	59.5	69.5	64.5	10
69.5-79.5	69.5	79.5	74.5	30
79.5-89.5	79.5	89.5	84.5	40
89.5-99.5	89.5	99.5	94.5	15

We also note the before and after class marks when plotting the graph. The before in this example is 44.5, while the after is 104.5. The score is shown on the x-axis, while the frequency is shown on the y-axis. Then the frequency polygon graph will look like the following:

Polygon graph



Example 2

Assume a class of 65 students has weights distributed as follows: 15 - 25, 25 - 35, 35 - 45, and 45 - 55. What is the point value for each weight category?

The formula used to obtain the class mark of a Frequency Polygon Graph is:

Classmark = (Upper Bound + Lower Bound) / 2

Therefore:

- Class interval 15-25 = (15 + 25)/2 = 20
- Class interval 25-35 = (25 + 35)/2 = 30
- Class interval 35-45 = (35 + 45)/2 = 40
- Class interval 45-55 = (45 + 55)/2 = 50

TESTS

There is data on the statistical value of Darmajaya Information System undergraduate students

67	44	45	78	67	78
83	72	66	45	45	73
56	62	74	79	36	67
67	91	56	94	48	65
45	34	67	67	68	55

Search for frequency distributionsDraw histograms and frequency polygons

TESTS

A travel agency recorded the number of passengers per week for the last 34 weeks as follows:

12, 45, 28, 42, 35, 15, 43, 24, 33, 32, 37, 27, 38, 57, 52, 48, 18, 26, 14,

54, 35, 36, 63, 22, 40, 34, 40, 55, 45, 34, 47, 37, 25, 68

Create a frequency distribution table of passengers per week with the following conditions:

- The lowest number for the first class is 10
- There are 6 classes
- Class interval is 10
- Determine the midpoint and relative frequency of each class
- Draw the histogram and polygon graph.