Sec “ 1 , 2 ”

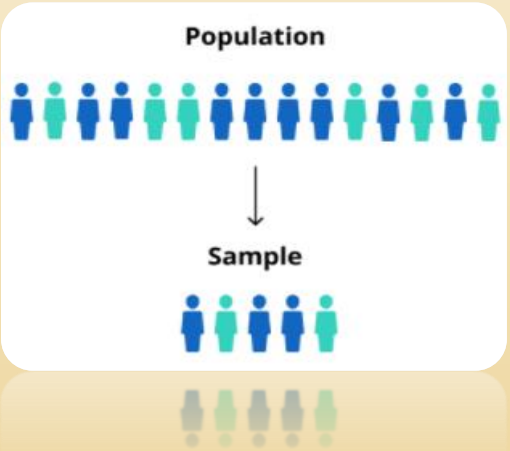
DESCRIBING DATA WITH TABLES AND GRAPHS

Statistics:

- Statistics is the study of the collection, analysis, interpretation, presentation, and organization of data. In other words, it is a mathematical discipline to collect, summarize data.

Samples and Populations:

The terms population and samples are defined in statistics as follows:



- **Population:** It is a collection of all possible individuals, about which we require information.
- **Sample:** A sample is a portion of the population of interest.

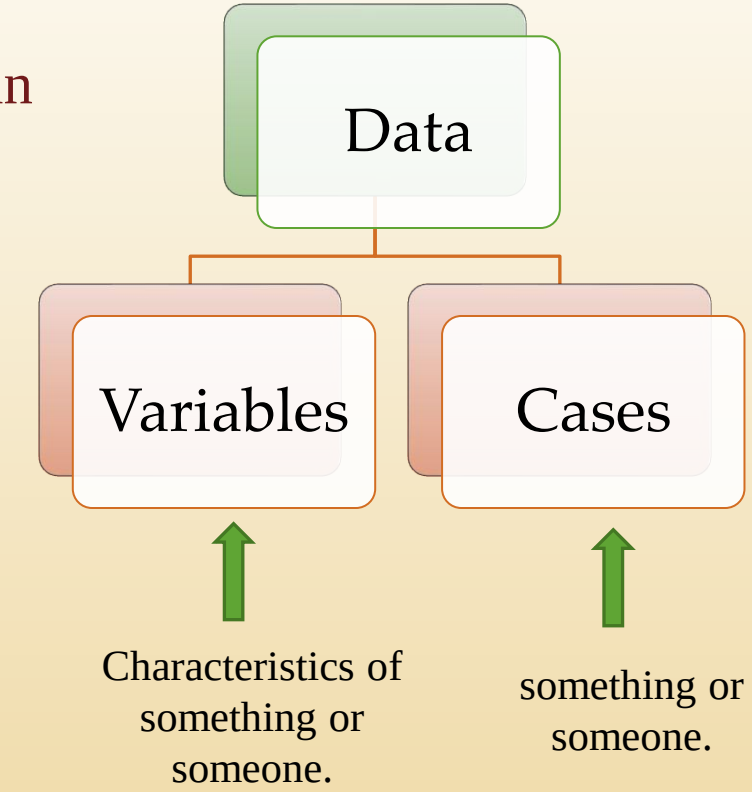
Random Sample:

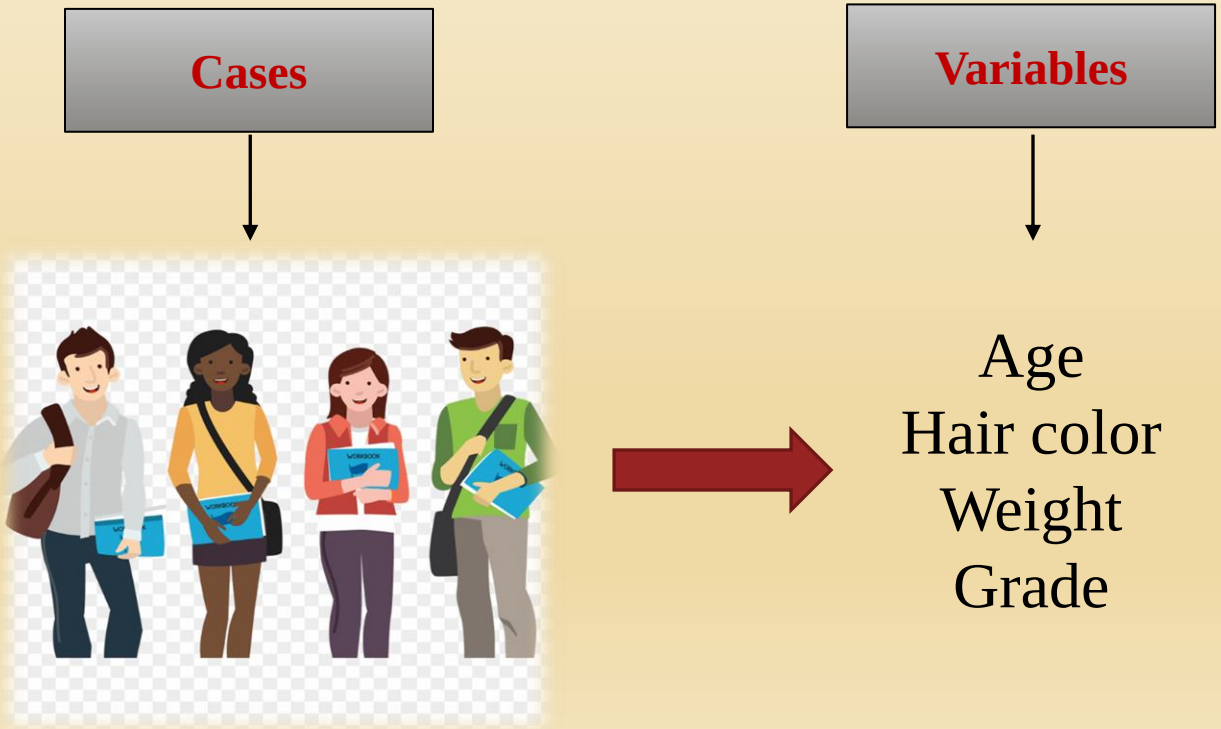
- A random sample of n observations $x_1, x_2, \dots, x_n$ is a sample that is chosen in such a way that each member of the population has the same chance of being selected for the sample.

Cases and Variables:

If you're conducting a study, you should think about your data in terms of cases and variables.

- **Cases:** Cases are the persons, animals, or things in your study.
- **Variables:** variables are the characteristics of interest.





Data Matrix:

Data matrix is the tabular format representation of cases and variables of your statistical study.

Each row of a data matrix represents a case and each column represent a variable.

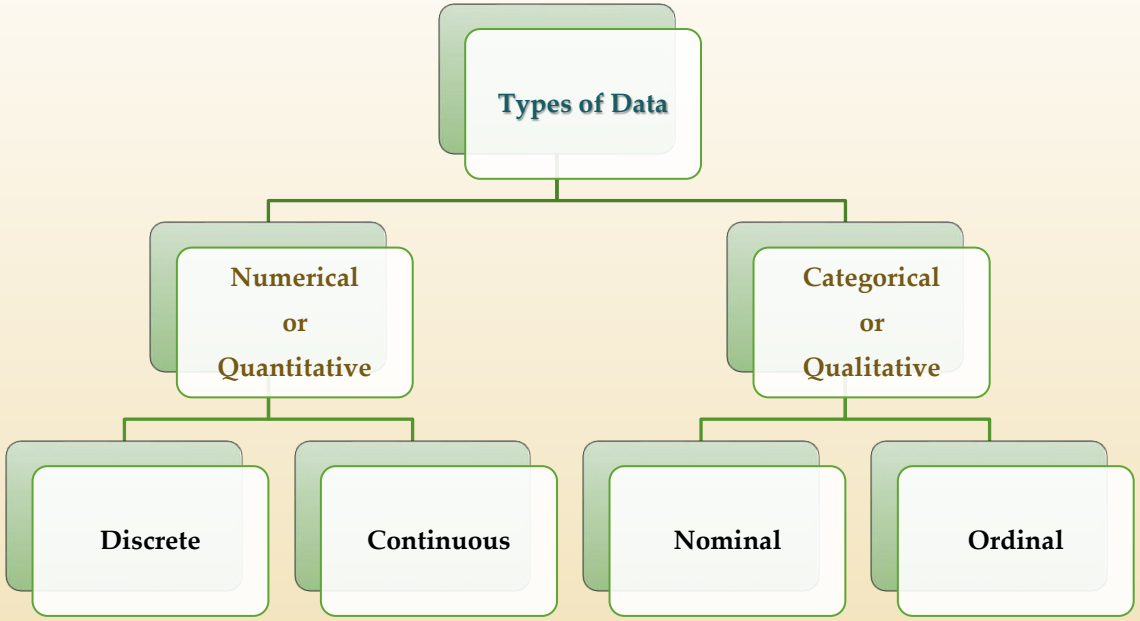
Variables

Cases

Name	Age	Hair Color	Weight	Grade
Ahmed	15	Black	54	90
Mona	13	Black	50	50
Sara	20	Brown	45	65
Ali	23	Blond	60	83

- **Numerical or Quantitative Data:**

Data can be measured, deals with numbers. Some examples of numerical data are height, length, size, weight, number of children in a family, minutes remaining in a class, and so on. The two different classifications of numerical data are:



(a) **Discrete Data :**

Discrete data can take only discrete values. Discrete information contains only a finite number of possible values.

Example: Number of students in the class (30, 35, 60, 42, …, etc.)

(b) **Continuous Data :**

Continuous data is data that can be calculated. It has an infinite number of probable values that can be selected within a given specific range.

Example: Temperature range (30 – 35)

- **Categorical or Qualitative Data:**

Data cannot be measured, describe characteristics. Some examples of categorical data are type of a car, eye color, hair color, hometown, person's gender, and so on. Sometimes categorical data can hold numerical values (quantitative value), but those values don't have mathematical sense (birthdate). The two different classifications of categorical data are:

(a) **Nominal Data :**

It cannot be ordered and measured.

Example: Names of people, gender, and nationality.

(b) **Ordinal Data :**

Ordinal data is a type of categorical data with an order. The variables in ordinal data are listed in an ordered manner.

Example: Ranking**,** and a position in class.

Graphic Presentation of Data

- The graphic presentation of data and information offers a quick and simple way of understanding the features and drawing comparisons. There are different types of graphical representation. Some of them are as follows:

[1] Dot – Plot (Dot – Chart) :

- A dot – plot is a graphical display of data using dots. It shows the frequency of data on a given number line. ‘ • ‘ is placed above a number line each time when that data occurs again.
- X – axis ⟶ what is being measured.

Example : How long does it take to eat breakfast

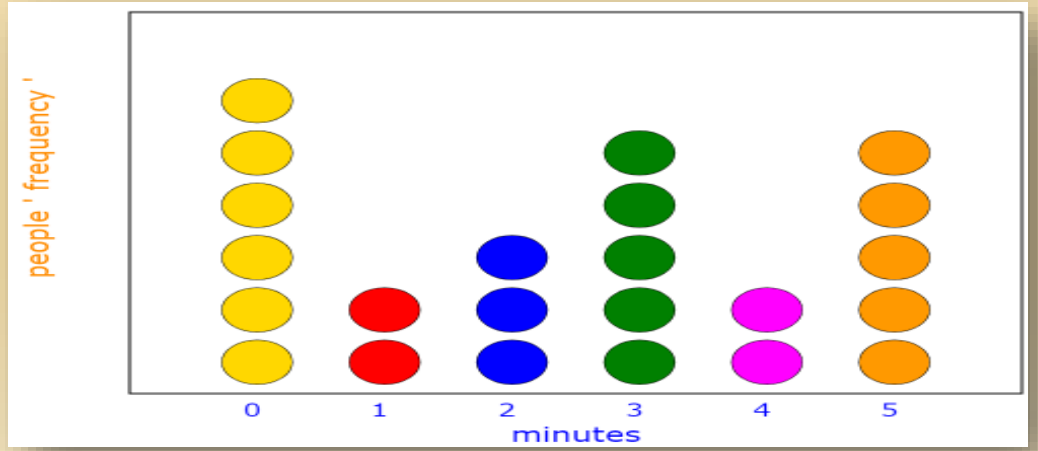
Minutes	0	1	2	3	4	5
people	6	2	3	5	2	5

(a) What is the sample size ?
(b) construct a dot – plot graph?

Answer

(a) Sample size = The sum of all people ‘ Total frequencies ’
= 6 + 2 + 3 + 5 + 2 + 5
= 23

(b)



Sheet (1)

1. The following measurements were recorded for the drying time, in hours, of a certain brand of latex paint.

3.4	2.5	4.8	2.9	3.6
2.8	3.3	5.6	3.7	2.8
4.4	4.0	5.2	3.0	4.8

5

3

Assume that the measurements are a simple random sample.

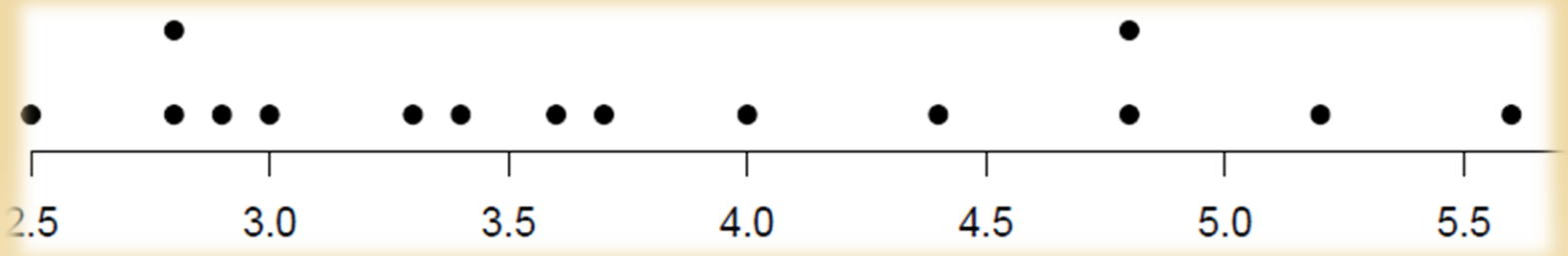
(a) What is the sample size for the above sample?

(b) Plot the data by way of a dot plot.

Answer

(a) Sample size $5 \times 3 = 15$.

(b)



[2] Bar – Graph (Bar – Chart) :

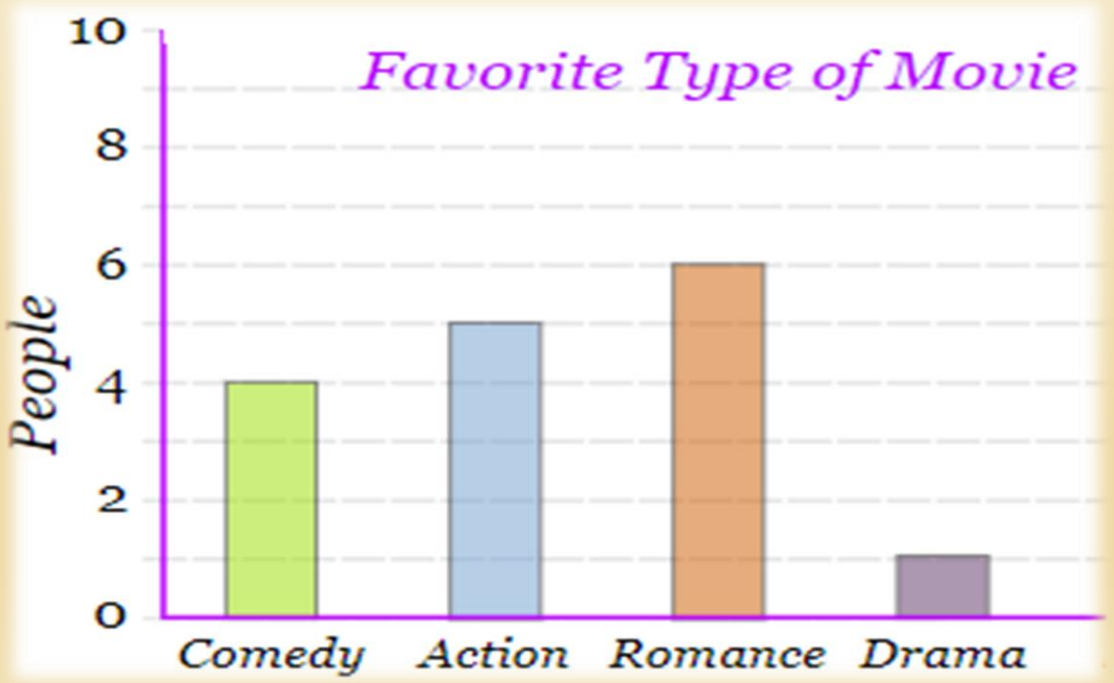
- A bar – graph is a graphical display of data using bars ‘ rectangles ’ of different heights. A bar – graph can be used to show something change over time or to compare items.
- X – axis ⟶ What is being measured.

  Y– axis ⟶ Frequency (the number for the amount of stuff being measured).
- There exist gaps between the bars.

Example (1) : Imagine you just did a survey of your friends to find which kind of movie they liked best ?

Movie	Comedy	Action	Romance	Drama
People	4	5	6	1

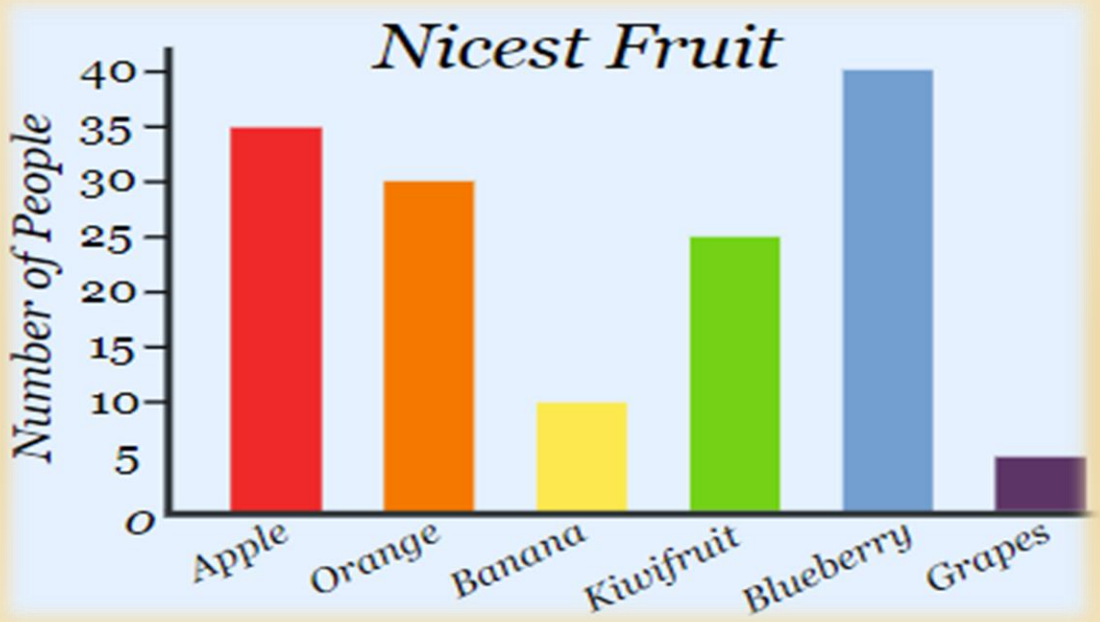
Plot the data by way of a bar – graph ?



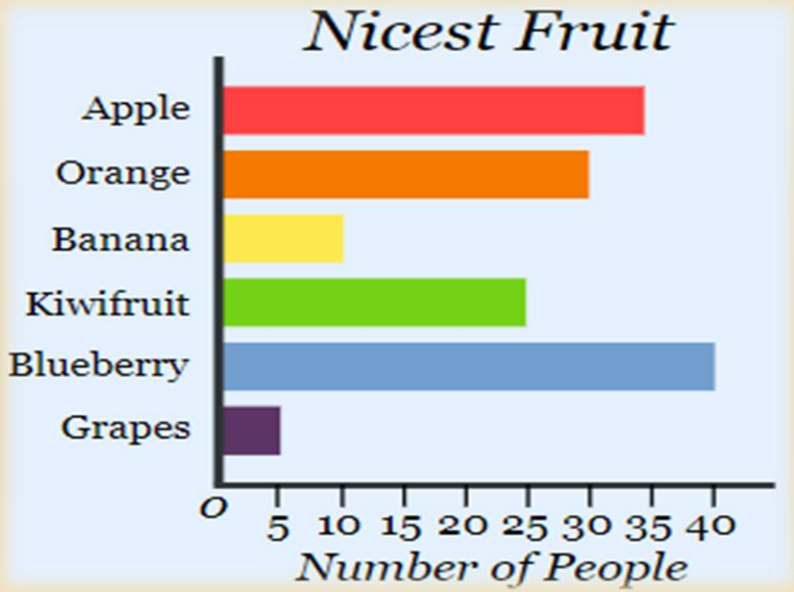
Example (2) : A survey of 145 people asked them "Which is the nicest fruit"?

Fruit	Apple	Orange	Banana	Kiwifruit	Blueberry	Grapes
People	35	30	10	25	40	5

Plot the data by way of a bar – graph ?



Bar Graphs can also be **Horizontal**, like this:



[3] Pie – Chart (Circle – Chart) :

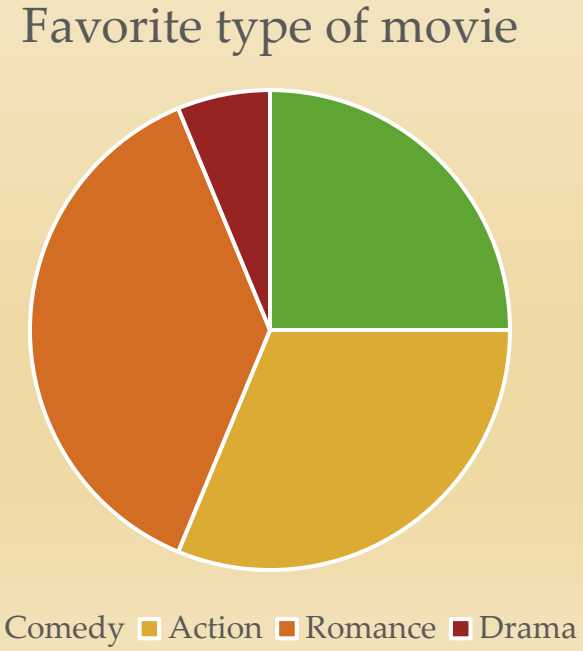
- a special chart that uses "pie slices" to show relative sizes of data.

Example (1) : Imagine you just did a survey of your friends to find which kind of movie they liked best ?

Movie	Comedy	Action	Romance	Drama
People	4	5	6	1

Plot the data by way of a Pie – chart ?

Movie	Comedy	Action	Romance	Drama	Total
People	4	5	6	1	16
Relative size	$\frac{4}{16} \times 100 = 25\%$	$\frac{5}{16} \times 100 = 31.25\%$	$\frac{6}{16} \times 100 = 37.5\%$	$\frac{1}{16} \times 100 = 6.25\%$	

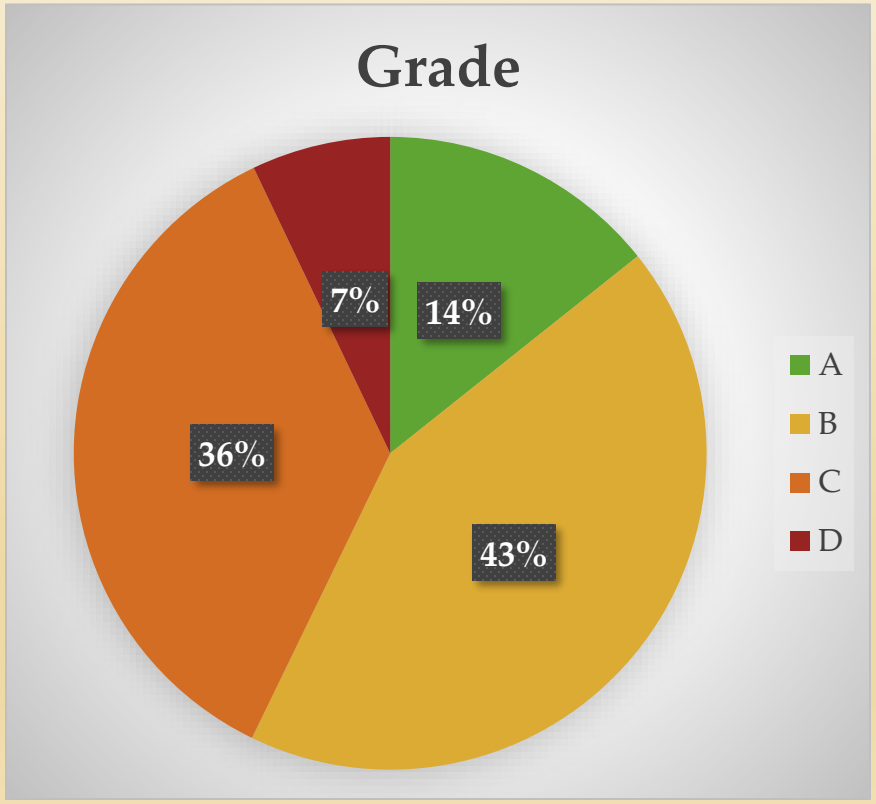


Example (2) : Here is how many students got each grade in the recent test:

A	B	C	D
4	12	10	2

Plot the data by way of a Pie – chart ?

A	B	C	D	Total
4	12	10	2	28
$\frac{4}{28} \times 100 = 14.3\%$	$\frac{12}{28} \times 100 = 42.9\%$	$\frac{10}{28} \times 100 = 35.7\%$	$\frac{2}{28} \times 100 = 7.1\%$	



Sheet (1)

7. According to the pie graph, which of the following statements is false ?

a) More than half the animals on the farm are cows.

True

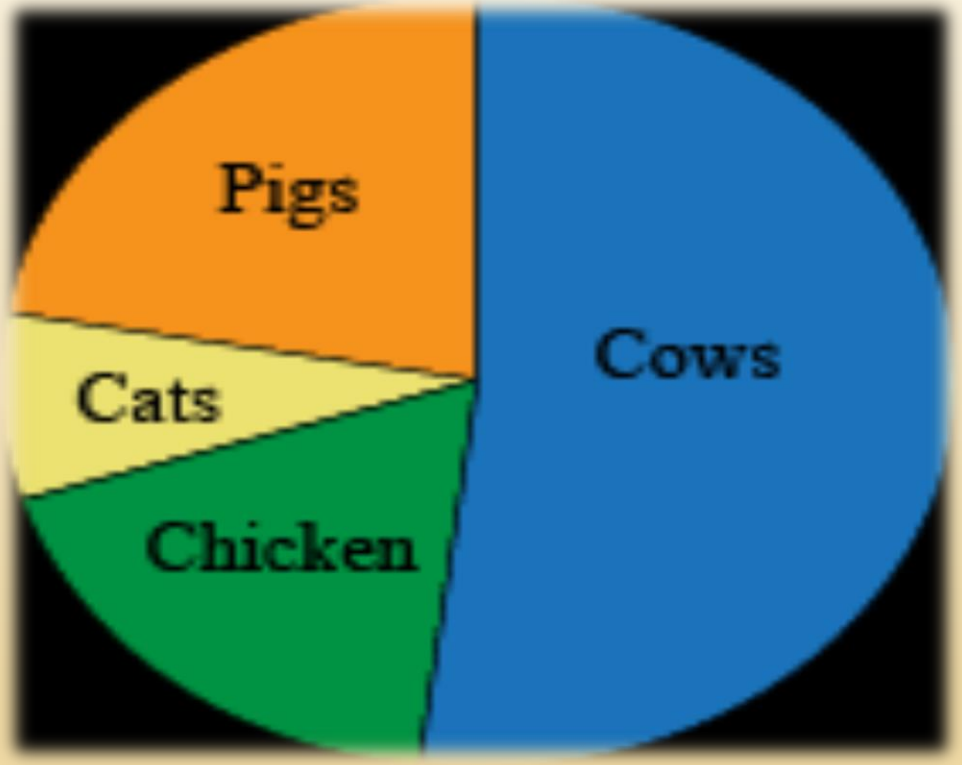
b) One quarter of the animals on the farm are chickens. False

c) There are more pigs than cats on the farm. True

d) Fewer than one quarter of the animals on the farm are pigs. True

e) No cats on the farm have given birth to cows.

True



[4] Histogram :

- A **histogram** is a graphical display of data using bars of different heights. In a **histogram**, each bar groups numbers into ranges. Taller bars show that more data falls in that range. A **histogram** displays the shape and spread of continuous sample data. In a histogram no gaps between the bars.

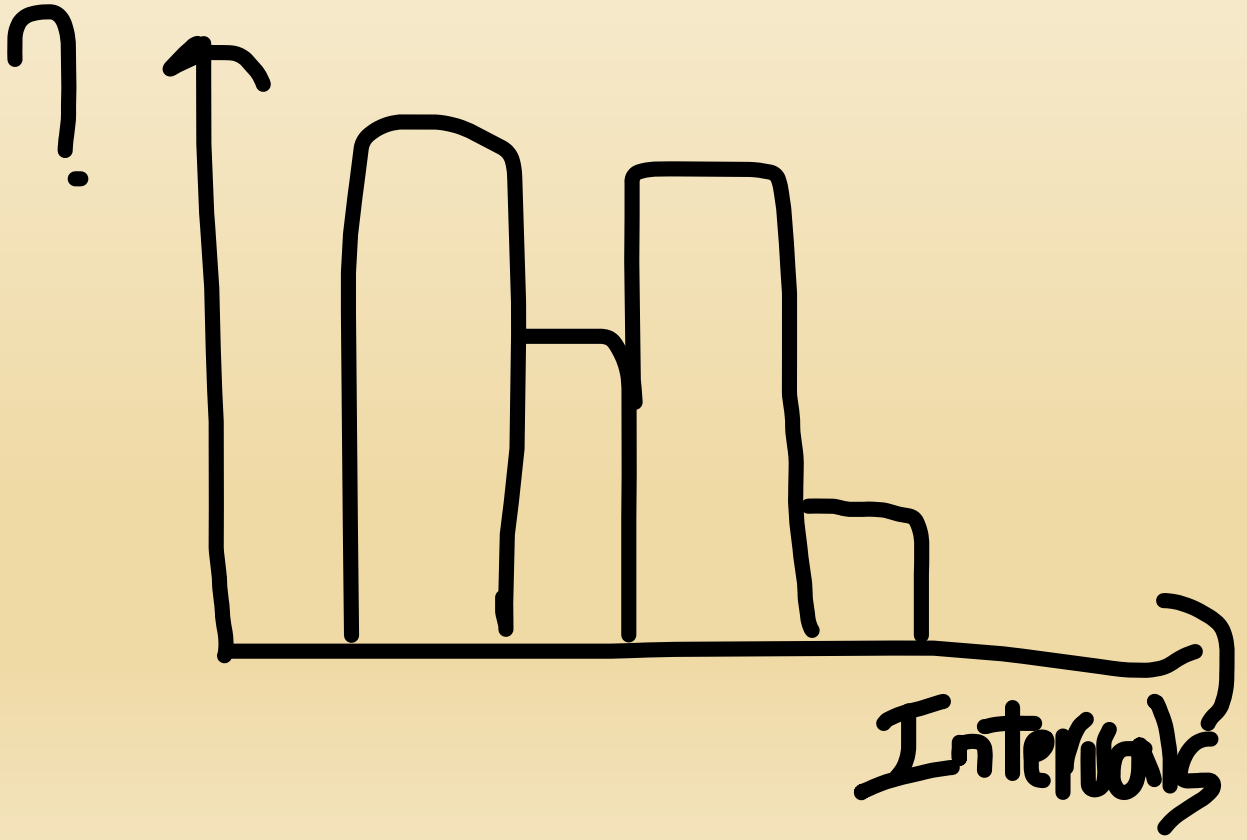
- **We have two cases:**

(1) Histogram for grouped data:

(a) Intervals ‘ classes ’ with equal width:

X – axis ⟶ Intervals.

Y – axis ⟶ Frequencies.



Example (1) : The following table shows a grouped frequency distribution for the statistics grades:

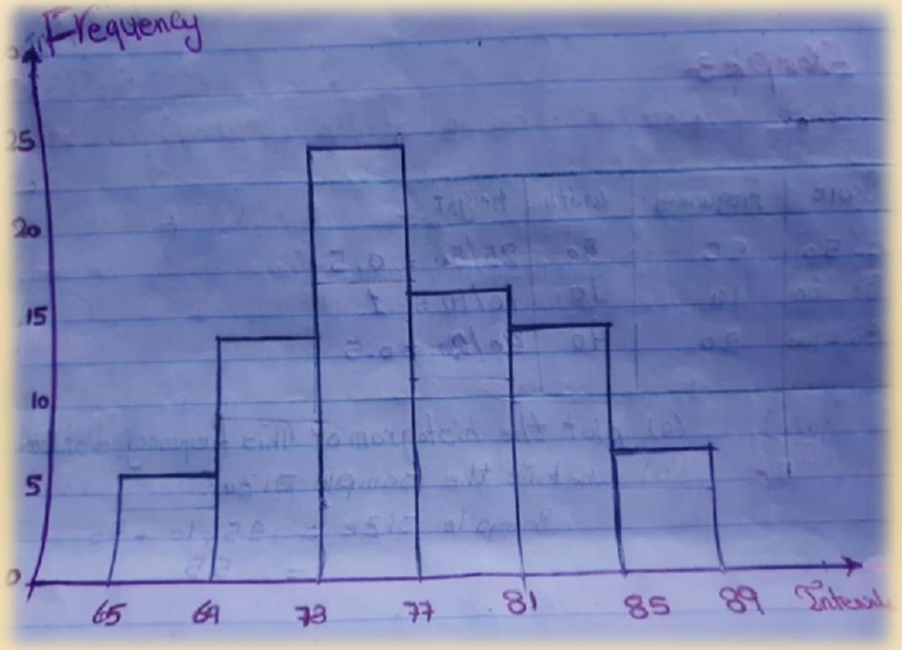
Intervals	Frequency	Width = upper limit – lower limit
65 – 69	6	4
69 – 73	13	4
73 – 77	24	4
77 – 81	16	4
81 – 85	14	4
85 – 89	7	4

a) What is the sample size ?

sample size = the sum of all frequencies

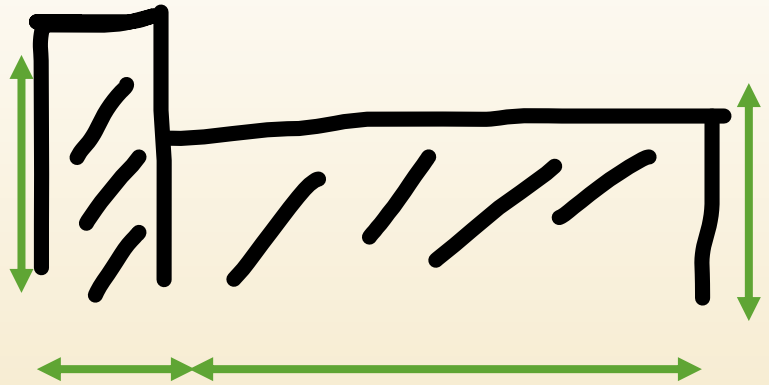
$= 6 + 13 + 24 + 16 + 14 + 7 = 80$

b) Plot the histogram of this frequency distribution table?



(1) Histogram for grouped data:

(b) Intervals ' classes ' with unequal width:



X – axis ⟶ Intervals.

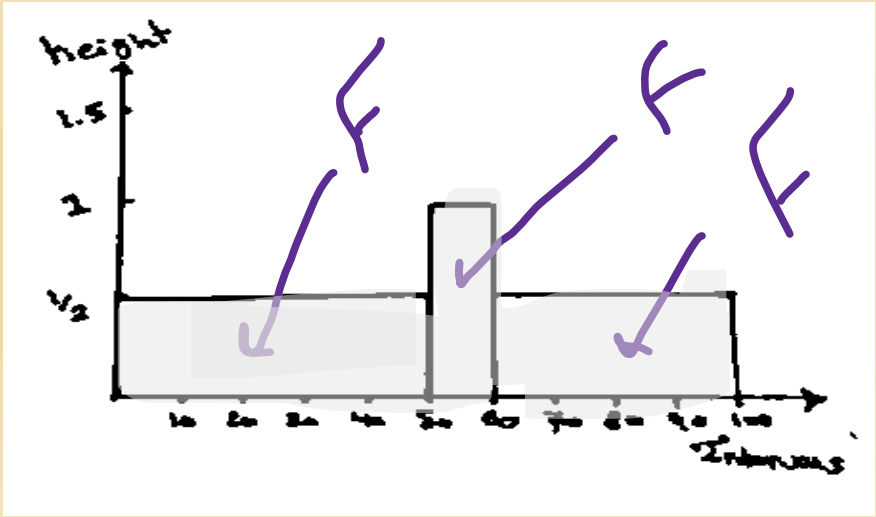
Frequency = Area of rectangle

= Height × Width ➡ $\text{Height} = \frac{Frequaency}{Width}$

Y – axis ⟶ Height.

Example (2) : The following table shows a grouped frequency distribution for the statistics grades:

Intervals	Frequency	Width = upper limit – lower limit	Height $= \frac{Frequaency}{Width}$
0 – 50	25	50	$^1/_2$
50 – 60	10	10	1
60 – 100	20	40	$^1/_2$



(2) Histogram for ungrouped data:

Step 1 : Find the smallest and largest data point.

Step 2 : Decide how many bins you need using

$\sqrt{n}$ **, where** n**: Sample size.**

let $k = \sqrt{n}$ **= number of intervals.**

Step 3 : Divided the range ‘ R ’of the data (R = largest value – smallest value) by the number of intervals ‘ k ‘ .

L = The width of each interval = $\frac{largest - smallest}{number\ of\ intervals} = \frac{R}{K}$

Step 4 : create the bin boundaries by starting with your smallest number ‘ or less ‘, and adding the bin size (L). The last interval should contain the largest number.

Sheet (1)

6. The following data represent the length of life in years, measured to the nearest tenth, of 30 similar fuel pumps:

5

6

2.0	3.0	0.3	3.3	1.3	0.4
0.2	6.0	5.5	6.5	0.2	2.3
1.5	4.0	5.9	1.8	4.7	0.7
4.5	0.3	1.5	0.5	2.5	5.0
1.0	6.0	5.6	6.0	1.2	0.2

Set up a frequency and a relative frequency distribution histogram.

Answer

(1) Largest value = 6.5 , Smallest value = 0.2

(2) n = Sample size = $6 \times 5 = 30$.

k = number of intervals = $\sqrt{n} = \sqrt{30} = 5.4 = 6$.

(3) L = Width of each interval = $\frac{largest - smallest}{k} = \frac{6.5-0.2}{6} = 1.05 = 1$.

0.2 — 1.2

1.2 — 2.2

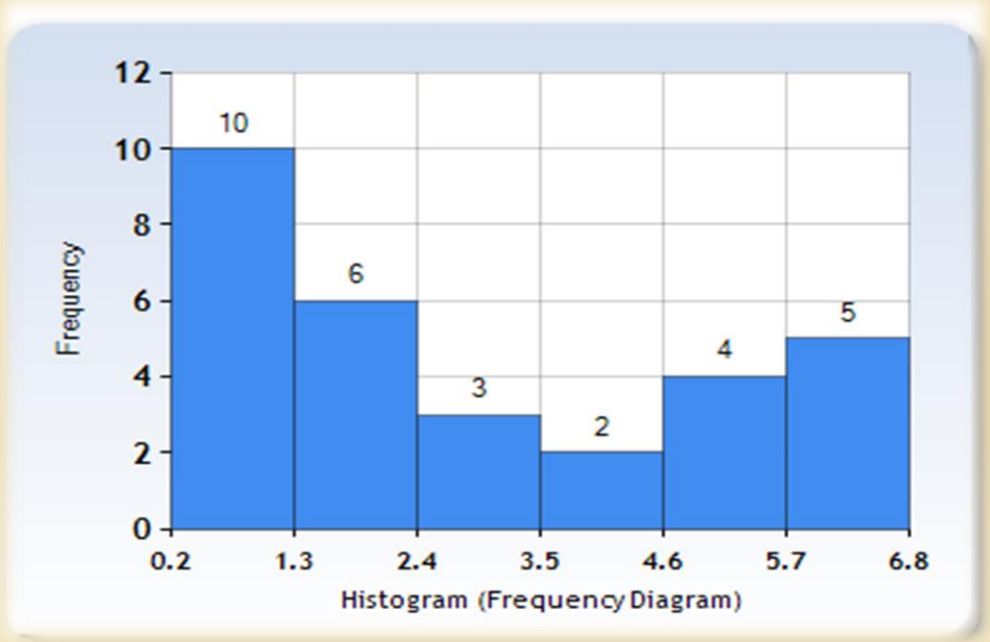
2.2 — 3.2

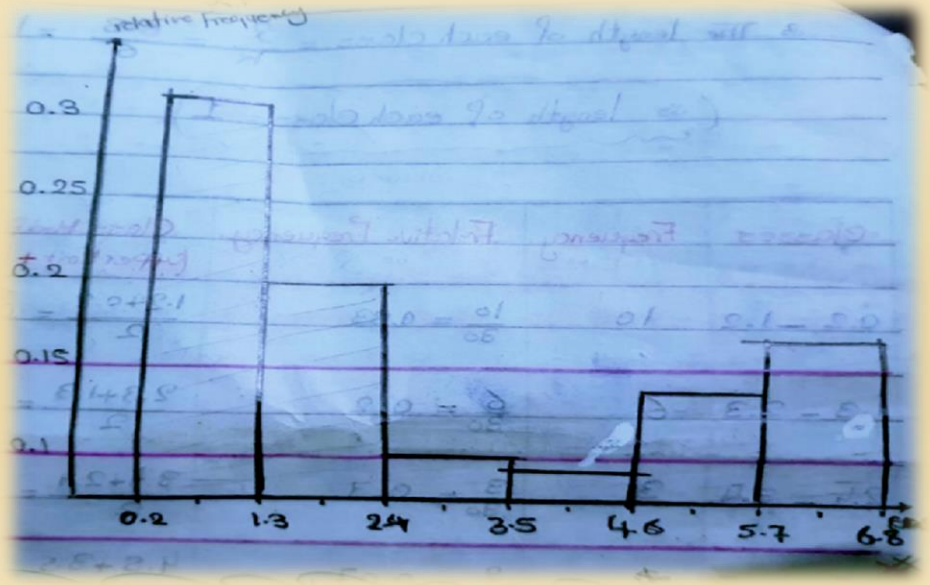
3.2 — 4.2

4.2 — 5.2

5.2 — 6.2

Intervals	Frequency ‘ F ’	Relative Frequency ‘ RF’ $= \frac{F}{n}$
0.2 – 1.2	10	$^{10}/_{30} = 0.33$
1.3 – 2.3	6	$^{6}/_{30} = 0.2$
2.4 – 3.4	3	$^{3}/_{30} = 0.1$
3.5 – 4.5	2	$^{2}/_{30} = 0.07$
4.6 – 5.6	4	$^{4}/_{30} = 0.13$
5.7 – 6.7	5	$^{5}/_{30} = 0.17$





Sheet (1)

7. The following data represent the length of life in seconds, of 50 fruit flies subject to a new spray in a controlled laboratory experiment:

17	20	10	9	23	13	12	19	18	24
12	14	6	9	13	6	7	10	13	7
16	18	8	13	3	32	9	7	10	11
13	7	18	7	10	4	27	19	16	8
7	10	5	14	15	10	9	6	7	15

5

10

Set up a frequency distribution histogram. Draw an estimate of the graph of the distribution?

Answer

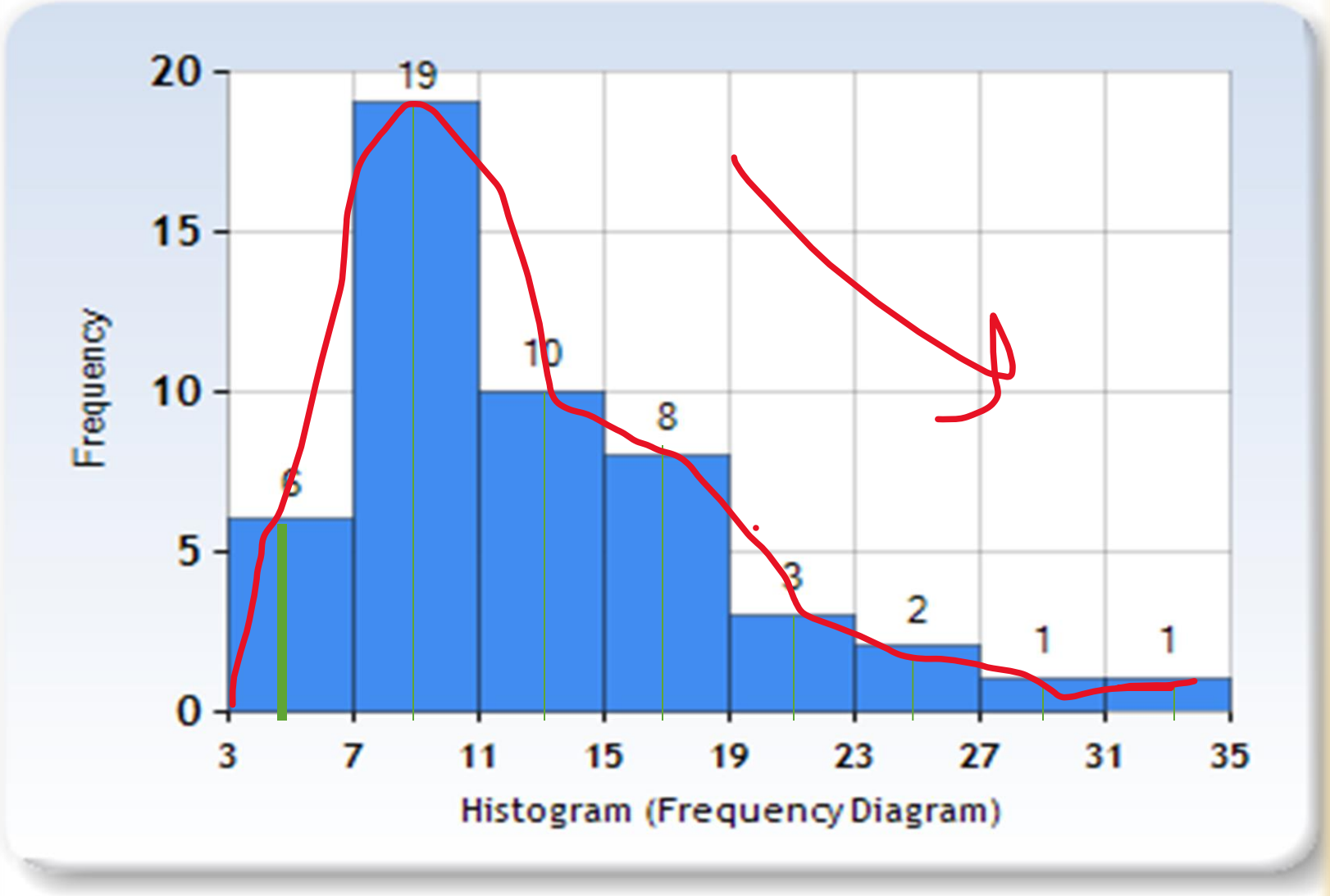
(1) Largest value $=\ 32\ $, Smallest value $= 3$

(2) $n =$ Sample size $= 10 \times 5 = 50.$

$k\ =$ number of intervals $= \sqrt{n} = \sqrt{50} = 7.07 = 8.$

(3) $L\ =$ Width of each interval $= \frac{largest\ -smallest}{k} = \frac{32-3}{8} = 3.6 = 4.$

Intervals	Frequency ‘ F ’	Class midpoint $= \frac{lower\ limit\ +upper\ limit}{2}$
3 – 7	6	$3+7/2 = 5$
7 – 11	19	$7+11/2 = 9$
11 – 15	10	13
15 – 19	8	17
19 – 23	3	21
23 – 27	2	25
27 – 31	1	29
31 – 35	1	33


20
15
10
5
0
Frequency
19
10
8
6
3
2
1
1
3
7
11
15
19
23
27
31
35
Histogram (Frequency Diagram)