Topic: Methods of Data collection-Primary and secondary Data

There are two types of data: 1. Primary Data and 2. Secondary Data

1. Primary Data: It is a term for data collected at source. This type of information is obtained directly from first hand sources by means of surveys, observations and experimentation and not subjected to any processing or manipulation and also called primary data. Primary data means original data that has been collected specially for the purpose in mind.It means someone collected the data from the original source first hand.

Primary data has not been published yet and is more reliable, authentic and objective. For example population census conducted by the government of India after every 10 years.

2. Secondary data: It refers to the data collected by someone other than the user i.e. the data is already available and analysed by someone else. Common sources of secondary data include various published or unpublished data, books, magazines, newspaper, trade journals etc.

COLLECTION OF PRIMARY DATA

Primary data is collected in the course of doing experimental or descriptive research by doing experiments, performing surveys or by observation or direct communication with respondents. Several methods for collecting primary data are given below-

1. Observation Method

It is commonly used in studies relating to behavioural science. Under this method observation becomes scientific tool and the method of data collection for the researcher, when it serves a formulated research purpose and is systematically planned and subjected to checks and controls.

(a) Structured (descriptive) and unstructured (exploratory) observation- When a observation is characterized by careful definition of units to be observed, style of observer, conditions of or observation and selection of pertinent data of observation it is a structured observation. When there characteristics are not thought of in advance or not present. it is a unstructured observation.

(b) Participant, Non-participant and disguised observation- When the observer observes by making himself more or less, the member of the group he is observing, it is participant observation but when the observer observes by detaching himself from the group under

observation it is non participant observation. If the observer observes in such manner that his presence is unknown to the people he is observing it is disguised observation.

(c) Controlled (laboratory) and uncontrolled(exploratory) observation- If the observation takes place in the natural setting it is a uncontrolled observation but when observation takes place according to some pre-arranged plans ,involving experimental procedure it is a controlled observation.

Advantages-

- Subjective bias is eliminated.
- Data is not affected by past behaviour or future intentions.
- Natural behaviour of the group can be recorded.

Limitations-

- Expensive methodology.
- Information provided is limited.
- Unforeseen factors may interfere with the observational task

2. INTERVIEW METHOD

This method of collecting data involves presentation of oral verbal stimuli and deeply in terms of oral- verbal responses. It can be achieved by two ways:-

(A)**Personal interview-** It requires a person known as interviewer to ask questions generally in a face to face contact to the other person. It can be –

Direct personal investigation- The interviewer has to collect the information personally from the services concerned.

Indirect oral examination- The interviewer has to cross examine other persons who are suppose to have a knowledge about the problem.

Structured interviews- Interviews involving the use of pre-determined questions and of highly standard techniques of recording

Unstructured interviews- It does not follow a system of pre-determined questions and is characteirzsed by flexibility of approach to questioning.

Focussed interview- It is meant to focus attention on the given experience of the respondent and its effect. The interviewer may ask questions in any manner or sequence with the aim to explore reasons and motives of the respondent.

Clinical interviews- It is concerned with broad underlying feeling and motives or individuals life experience which are used as method to collect information under this method at the interviewer direction.

Non directive interview- The interviewer`s function is to encourage the respendent to talk about the given topic with a bare minimum of direct questioning.

Advantages-

- More information and in depth can be obtained.
- Samples can be controlled.
- There is greater flexibility under this method
- Personal information can as well be obtained.
- Mis-interpretation can be avoided by unstructured interview.

Limitations

- It is an expensive method.
- More time consuming.
- Possibility of imaginary info and less frank responses.
- High skilled interviewer is required

(B) **Telephonic interviews-** It requires the interviewer to collect information by contacting respondents on telephone and asking questions or opinions orally.

2. QUESTIONNAIRE

In this method a ouestionnaire is sent (mailed) to the concerned respondents who are expected to read, understand and reply on their own and return the questionnaire. It consists of a number of questions printed or typed in a definite order on a form or set of forms.

It is advisable to conduct a 'pilot study' which is the rehearsal of the main survey by experts for testing the questionnaire for weaknesses of the questions and techniques used.

Essential of a good questionnaire

It should be short and simple.

- Questions should processed in a logical sequence.
- Technical terms and vauge expressions must be avoided.

- Control questions to check the reliability of the respondent must be present.
- Adequate space for answers must be provided.
- Brief directions with regard to filling up of questionnaire must be provided.
- The physical appearances-quality of paper, colour etc must be good to attract the attention of the respondent

Advantages

- Free from bias of interviewer.
- Respondents have adequate time to give answers
- Respondents are easily and conveniently approachable Large samples can be used to be more reliable.

LIMITATIONS

- Low rate of return of duly filled questionnaire.
- Control over questions is lost once it is sent.
- It is inflexible once it is sent.
- Possiblitty of ambiguous omission of replies.
- Time taking and slow process.

3. SCHEDULES

This method of data collection is similar to questionnaire method with difference that schedule are being filled by the enumerations specially appointed for the purpose. Enumerations explain the aims and objects of the investigation and may remove any misunderstanding and help the respondents to record answer. Enumerations should be well trained to perform their job,he/she should be honest hardworking and patient. This type of data is helpful in extensive enquiries however it is very expensive.

Collection of secondary data

A researcher can obtain secondary data from various sources.Secondary data may either be published data or unpublished data.

Published data are available in:

a. Publications of government.

b. Technical and trade journals.

c. Reports of various businesses, banks etc.

d. Public records.

e. Stastistical or historical documents.

Unpublished data may be found in letters, diaries, unpublished biographies or work. Before using secondary data it must be checked for the following characteristics-

1. Reliability of data- Who collected the data? From what source? Which method? Time? Possibility of bias? Accuracy?

2. Suitability of data- The object scope and nature of the original enquiry must be studies and then carefully scrutinize the data for suitability.

3. Adequency- The data is considered inadequate if the level of accuracy achieved in data is found inadequate or if they are related to an area which may be either narrower or wider than the area of the present enquiry.

Collection and Representation of Data

Collection of data : Primary and secondary data. Specific aspects of statistical data. Statistical units.

Classification of data : Meaning of classification. What is the need of classification? Objectives of classification. Types of classification. Classification according to attributes. Classification according to class intervals. Construction of frequency distribution – Discrete frequency distribution, Continuous frequency distribution. Number of classes and their magnitudes.

Tabulation of data : Need of tabulation. Significance of tabulation. Rules of tabulation, Types of tables

Graphic presentation of data : Types of graphs. Methods of preparation of graph. Histogram. Frequency polygon. Frequency curve. Relative frequency map. Cumulative frequency curve or ogive. Dot diagram. Significance of graphic representation. Limitations of graphic presentation.

Diagrammatic representation of data : Line diagram. Bar diagram. Pie diagram. Exercise.

Collection of Data

The study of techniques of collection of data and its presentation enables one to draw reliable conclusions from the collected data, which is obtained through various experiments and helps in analysis and interpretation.

Primary and secondary data

Data obtained by an investigator for the first time are original one and are alled as **primary data**. Whereas if data are collected by some other persons

BC-33)

and have been processed at least once then such data are known as secondary data. In scientific researches only primary data are used.

Specific aspects of statistical data

There are four specific aspects of statistical data.

1. **Collection of data.** The first step in a statistical investigation is collection of data. A data collected in the original form is called **raw data** or ungrouped data. The data obtained by personal investigation is called primary data. There are several methods of collection of primary data. In scientific research, data is obtained from experimental results.

2. **Presentation of data.** Collected data are presented in an orderly manner to facilitate statistical analysis. There are different methods of presentation of data such as tables, diagrams, graphs etc.

3. **Analysis.** Data represented in tables, diagrams or graphs are analysed carefully. There are numerous methods of analysis of presented data. Measures of central tendency, measures of dispersion, correlation, regression etc. are a few examples of methods of analysis of presented data.

4. **Interpretation.** Drawing conclusion from analysis of data is called interpretation. Correct interpretation leads to valid conclusion.

Statistical units

The units in which the measurements are made in any statistical investigation are called the statistical unit. In life science survey of the unit may be a species or an individual.

Classification of Data

Data obtained from an experiment are classified i.e. converted into frequency distribution to make things simple and compact.

Meaning of classification

Classification is a process of condensation of raw data into systematized data that can be put up to a more systematic and proper use.

Need of classification

Statistical calculations from raw data are not advisable because it will require too much of time, space and labour. It is better to summarize the raw data into a frequency distribution table and then to give statistical treatment. The purpose of classification of data is to organise the data into a more compact form without obscuring the essential information contained in the values. The frequency distribution presents data very concisely indicating the number of repetition of values of variables. It records how frequently a variable occurs in

Objectives of classification

Process of classification is carried out with the following objectives—

(1) To bring out the unity of attributes out of the diversified things in collected data.

(2) To condense the universe and to make things easily intelligible.

(3) To make the study and comparison easier.

(4) To give prominence to the important information gathered while dropping out unnecessary elements.

(5) To put up the collected material to statistical treatment.

(6) To help the drafting of the final report.

(7) To simplify the complexities of the raw data and make it possible to draw statistical inferences.

(8) To make proper use of the collected data.

Types of classification

The type of classification to be carried out depends on the types of data being dealt with. In biostatistics we usually meet two types of classification—

(1) Classification according to attribute and

(2) Classification according to class intervals.

1. Classification according to attributes. Various enquiries deal with phenomena which can be reduced to quantitative characteristics and also with those which cannot be reduced to quantitative characteristics. In this case grouping is done according to the attribute that a particular phenomenon possesses. For example, male/female, intelligent/dull, red/white etc. Such a classification is known as classification according to attributes. This may be simple or manifold.

(a) Simple classification. Here the object of classification can be put under two columns. For example, human population may be classified into two categories as males and females, married and unmarried, literate and illiterate, rich and poor etc.

(b) Manifold classification. It is the classification in which more than one attribute is studied simultaneously. For a fish population, for an example, we may first divide the population into male and female on the basis of the attribute "sex". Each of these classes may be further subdivided into healthy and diseased on the basis of attribute "health". The subdivision may further be divided on the basis of other attributes.

2. Classification according to class interval. When classification of observations is done according to some measurable quantity, such a classification is known as quantitative classification or numerical classification or frequency distribution. In this type of classification direct quantitative

measurement of data is possible. Every frequency distribution possesses two characteristics—

(1) The data are simultaneously collected at a point in time so that the time element is not variable.

(2) The classification is made according to the magnitude of the variable rather than its quantitative or geographical characteristics.

Construction of frequency distribution

If there are repetitions in individual values or items of investigation suitable frequency table can be framed. These frequency tables may be discrete or continuous in nature, but they must maintain the frequency concerned in their respective classes.

Discrete frequency distribution. All the observations are listed in ascending or descending orders.

Following raw data were obtained in a biological experiment. Rate of reproduction (Fecundity) of 50 fishes was recorded as follows —

Raw Data (A)

80	70	70	70	16	50	20	20	20
45	16	50	30	65	40	30	50	50
70	45	20	70	2	79	16	20	19
40	50	30	2	45	30	50	45	30
40	45	80	50	39	50	50	20	30

A frequency distribution table is framed on the basis of above raw data. Following steps are taken while framing frequency distribution table.

Converting raw data in arrayed data. The primary duty of a biostatistician is to convert **raw data** in **arrayed data**. This can be done by arranging the raw data into ascending or descending orders. For biostatistics data are usually arranged in an ascending order. The above **raw data** arranged in **ascending order** to make **arrayed data**

Arrayed Data (B)

2	2	2	16	16	16	16	20	20	20	20	20	20	20	30
30	30	30	30	30	30	30	40	40	40	45	45	45	45	45
45	50	50	50	50	50	50	50	50	50	65	65	70	70	70
70	70	80	80											

Framing a simple frequency table (grouped series in discrete condition) :

(1) A table of two columns is framed. First column contains **variables** and second column contains **repetition number** i.e. frequency of variables.

(2) On persual of the above arrayed data (B), we find that variable 2 is obtained thrice. Therefore, frequency 3 is mentioned in second column i.e. in the frequency column. Variable 16 is obtained four times and hence 4

is mentioned. In the same ... variables ... mentioned and following frequency distribution table is framed —

Table 4.1

Variable	2	16	20	30	39	40	45	49	50	65	70	79	...
Frequency	3	4	7	7	1	3	5	1	9	2	5	1	2

Framing a continuous distribution table. When the number ... observations are very large, the discrete frequency distribution consumes a ... of space and time. So we group the data in some predetermined intervals ... groups. Such a classification is known as classification according to cla... intervals and represents the continuous distribution. The classification of da... according to the class intervals involves the following three basic problems :

(1) Determining the number of classes and their magnitudes.

(2) Choosing the appropriate class limits.

(3) Counting the number in each class.

Number of classes and their magnitude. In a continuous frequency distribution, the number of classes into which the observations are divided plays a very important role. The larger the number of observations, more are the number of classes. Though the number of classes in a continuous frequency distribution is not fixed, it should not be too many or too less. An ideal number of classes for many frequency distribution would be that which gives the maximum information in the clearest fashion. In practice, we usually take a minimum of 3 and maximum of 20 classes.

Thus, if the highest value in an investigation be 76 and the smallest value 3, the difference would be 76 –3=73. Suppose, we would like to divide the information into 10 classes, then the class number would be 73/10 =7.3.

Number of classes can be decided with the help of **Sturge rule**. This rule tells that the number of classes would be —

$$k = 1 + 3.322 \log N$$

Where, N = Total number of observations.

log = logarithm of the number

For example, if 100 observations are being made, the number of classes will be :

$$k = 1 + (3.322 \times 2) = 1 + 6.644 = 7.644 \text{ or } 8.$$

Classification of arrayed data into continuous frequency distribution. Values of variable are kept into ordered class intervals. The width or range of class is called class interval. The width of class intervals is kept at a uniform size and is denoted by *i*.

Size of class interval depends on the range of data and the number of classes. The range is the difference between the highest and lowest value of the

variable. The class interval would be equal to the difference between the highest and the lowest values of the variable divided by the number of classes. The following formula may be used to estimate the class interval :

$$i = \frac{H - L}{K}$$

where, i = width of class interval : H = highest value of variable
L = lowest value of variable k = the number of classes.

To illustrate the construction of a frequency distribution table in class interval, let us consider the arrayed data (B) which represents the rate of reproduction in 50 fishes of a species.

First of all we have to find out the number of classes.

According to **sturge rule** $k = 1 + 3.322 \log N$.

Here, $k = 1 + (3.322 \times 1) = 1 + 3.322 = 4.322$ or 5.

Therefore, the mumber of classes may be 5.

After deciding the number of classes, we should find out the suitable width of class interval.

class interval or $i = \frac{H - L}{K}$

See **arrayed data (B)**. It shows the highest value of observation is 80 and lowest value is 2 and number of classes i.e. k is 5.

$$i = \frac{80 - 2}{5} = \frac{78}{5} = 15$$

Thus, a table may be framed having width of class interval 10 or 15. But for convenience, an investigator can take suitable number of classes and width of class interval. Frequency distribution table in class interval may be prepared in two ways :

(a) Overlapping frequency table or exclusive method. Values of variables are grouped in such a fashion that the upper limit of one class interval is the lower limit of succeeding class interval. An overlapping class interval frequency distribution table can be prepared using data of Table 4.1.

Data of table 4.1 tells that the rate of reproduction of the given species of fish ranges between 2 and 80. We can keep the width of class interval 10. Then the range of first class interval will be 0 –10, 2nd between 10 –20, 3rd between 20 –30 and so on. Here one thing is **remarkable** — fishes having rate of reproduction upto 9 are taken into consideration in the first class interval. On persual of table 4.1, it appears that 3 fishes come under this class interval. Therefore, the frequency of class interval 0 –10 is 3. Fishes having 10 rate of reproduction have to be included in the succeeding class interval. Four fishes come under second class interval. Hence, frequency of 2nd class interval is 4 (Table 4.2).

(b) ... variables are grouped ... overlapping frequency ... do not overlap the succeeding class ... 4.3 can be prepared using the data of table 4.1.

Here class interval may be 1 – 10, 11 – 20, 21 – 30 and so on keeping width of class interval 10. Here upper limit of one class interval is not overlapped by lower limit of preceeding class interval (Table 4.3).

Table 4.2

Overlapping frequency table	
Class interval	Frequency
0 – 10	3
10 – 20	4
20 – 30	7
30 – 40	8
40 – 50	9
50 – 60	9
60 – 70	2
70 – 80	6
80 – 90	2
	$\Sigma f = 50$

Table 4.3

Non-overlapping frequency table	
Class interval	Frequency
1 – 10	3
11 – 20	11
21 – 30	7
31 – 40	4
41 – 50	15
51 – 60	0
61 – 70	7
71 – 80	3
81 – 90	0
	$\Sigma f = 50$

Note – For biostatistics usually non-overlapping frequency distribution table is used.

Preparation of cumulative frequency distribution table, relative cumulative frequency table and % cumulative frequency table

Following steps have to be taken to frame cumulative frequency table, relative cumulative frequency table and % cumulative frequency table —

A table of five columns is framed. In first column **class interval** is given. In 2nd column **frequency** is noted. In 3rd column **cumulative frequency** are mentioned. (For the preparation of cumulative frequency distribution, add frequency of first class interval with the frequency of 2nd class interval. Say the frequency of first class interval is 3, then the frequency of first class interval is noted as 3. The frequency of second class interval is 11 then add 11 with 3. Total comes to 14. This 14 is cumulative frequency of 2nd class interval. Now if the frequency of 3rd class interval is 7 then add 14 and 7. This comes to 21. Frequency of 3rd class interval is mentioned as 21. In the same fashion cumulative frequency of each class interval is ascertained). Fourth column is for **relative cumulative frequency**. (For preparation of relative

Tabulation of Data

Tabulation is a process of orderly arrangement of data into series of rows and columns where they can be read in two dimensions. Tabulated data can be more easily understood than those which are not tabulated. A tabular arrangement makes the summation of items and detection of errors and omissions easier.

Need of tabulation

Statistical tabulation is one of the simplest and most revealing devices meant for summarising data in an orderly manner so as to bring out its essential features and chief characteristics. A table is a systematic arrangement of statistical data in columns and rows. Tables make it possible for the analyst to present a huge mass of data in a detailed orderly manner within the minimum space. Tabular presentation of raw data is the cornerstone of statistical reporting.

Significance of tabulation

It simplifies complex raw data. It facilitates comparison between different values of a parameter. It gives identity to the data. It reveals patterns within the figures. It facilitates the detection of errors and omissions.

Rules of tabulation

There is no hard and fast rule for tabulation of data because much depends upon the given data and requirement of the survey. However, following general considerations may be kept in view while tabulating data —

(1) The table should suit the size of the paper usually with more rows than columns. It is desirable to make a rough draft of the table before the figures are entered into it.

(2) In all tables; number, heading and sub-heading should be arranged in some systematic order such as alphabetical, chronological, attributes etc.

(3) The points of measurement should be clearly defined and given in the table such as length in cm or weight in g etc.

(4) Figures should be rounded off to avoid unnecessary details in the table and a footnote to this effect should be given.

(5) The table should not be overloaded with details, rather a number of tables should be prepared. Each table should be complete in itself and serve a particular purpose.

Tabulation of Data

Tabulation is a process of orderly arrangement of data into series of rows and columns where they can be read in two dimensions. Tabulated data can be more easily understood than those which are not tabulated. A tabular arrangement makes the summation of items and detection of errors and omissions easier.

Need of tabulation

Statistical tabulation is one of the simplest and most revealing devices meant for summarising data in an orderly manner so as to bring out its essential features and chief characteristics. A table is a systematic arrangement of statistical data in columns and rows. Tables make it possible for the analyst to present a huge mass of data in a detailed orderly manner within the minimum space. Tabular presentation of raw data is the cornerstone of statistical reporting.

Significance of tabulation

It simplifies complex raw data. It facilitates comparison between different values of a parameter. It gives identity to the data. It reveals patterns within the figures. It facilitates the detection of errors and omissions.

Rules of tabulation

There is no hard and fast rule for tabulation of data because much depends upon the given data and requirement of the survey. However, following general considerations may be kept in view while tabulating data —

(1) The table should suit the size of the paper usually with more rows than columns. It is desirable to make a rough draft of the table before the figures are entered into it.

(2) In all tables; number, heading and sub-heading should be arranged in some systematic order such as alphabetical, chronological, attributes etc.

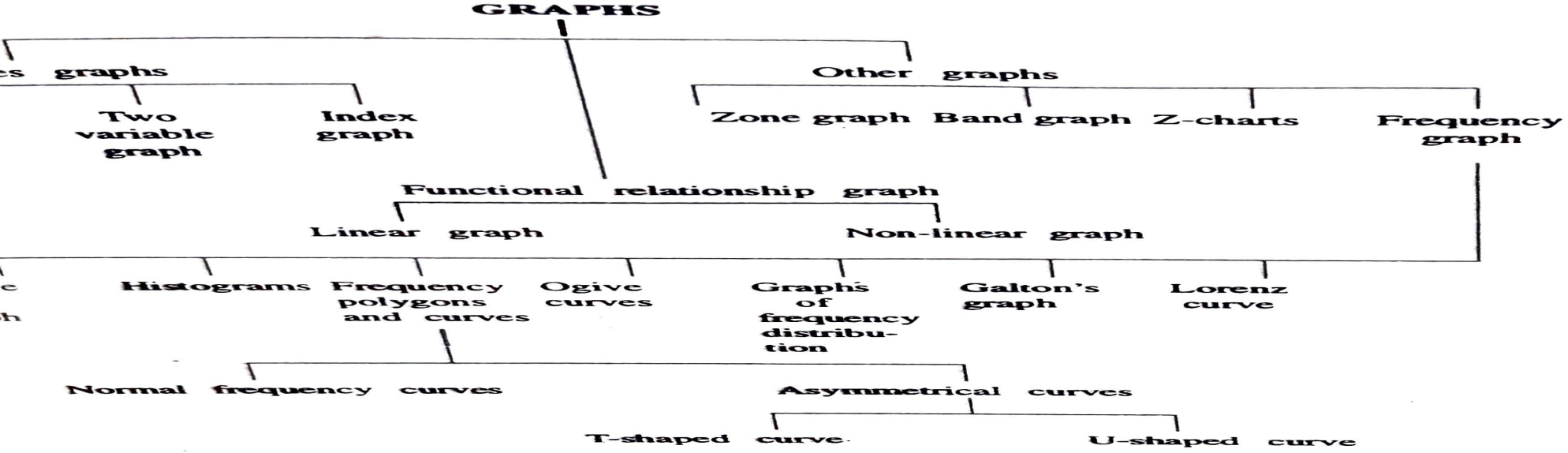
(3) The points of measurement should be clearly defined and given in the table such as length in cm or weight in g etc.

(4) Figures should be rounded off to avoid unnecessary details in the table and a footnote to this effect should be given.

(5) The table should not be overloaded with details, rather a number of tables should be prepared. Each table should be complete in itself and serve a particular purpose.

Representation of Data

resentation of data is widely used to present data in a simple, clear e manner. A graph is a visual form of presentation of data where s can be made between two or more phenomena. It is rightly said ring of a line is more powerful in its effect in mind than a tabulated Different types of graphs are presented as follows —



Time series graphs

It shows changes in values of a variable over passage of time. The graph of a time series for continuous data is referred to as histogram.

Functional relationship graphs

Sometimes two variables under consideration are such that one is completely dependent on the other. Two types of functional relationship graphs are important : (1) Linear curve and (2) Non-linear curve.

(1) **Linear curves** are those which always give a straight line when plotted on a graph paper.

(2) **Non-linear curves** are those which give a non-linear line when plotted on a graph.

Other graphs

Four important types of graphs can be studied under this heading.

(1) **Zone graphs** are plotted when the differences between maximum and minimum values of a variable have to be emphasized.

(2) **Band graphs** are plotted when it becomes necessary to represent the total and its components for successive time periods.

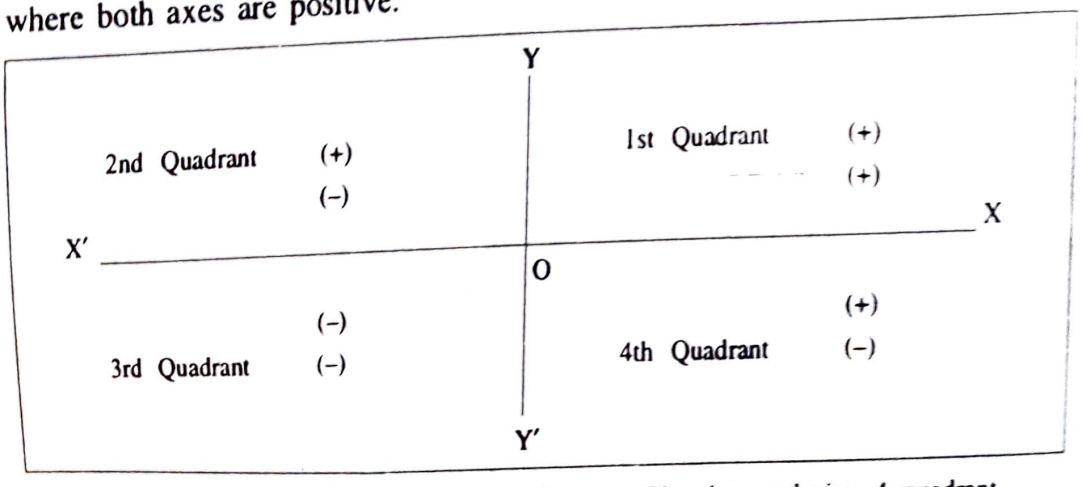
(3) **Z-chart** is also known as zee chart. It consists of three curves on the same axis.

(4) **Frequency graph.** Frequency distributions of all types are represented by means of graphs precisely for the same reasons for which graphs are plotted for other types of data. Such graphs are known as frequency graphs.

Methods of preparation of graph

Some basic knowledge is essential to prepare frequency graph. It is prepared with the help of two lines. The horizontal line is called Abscissa *i.e* X-axis representing independent variable and the vertical line called Ordinate i.e. Y-axis representing dependent variable. The meeting point of X and Y axis is called Zero (0) or origin point.

The right part of **'X' axis** from the zero point (0) is positive (+) and left part is negative (–). Likewise the upper part of **'Y' axis** from zero point is positive while the lower part is negative. 'X' and 'Y' axis intersect each other at '0' point and graph is divided into 4 parts. Each part is called Quadrant. Upper right part is called first Quadrant where 'X' and 'Y' both axis are positive. Upper left part is called second Quadrant. Here 'X' axis is negative (–) and 'Y' axis is positive (+). The lower left part is called third Quadrant where both 'X' and 'Y' axis is negative. The lower right part is known as fourth Quadrant where 'X' axis is positive (+) and 'Y' axis is negative (–). Mostly first quadrant is used for graphical representation of statistical data. where both axes are positive.



Two axes 'X' and 'Y' intersecting each other on 'O' point producing 4 quadrant.

Units of representation

Appropriate unit bar line is required to present the statistical data in graph. Suppose, one has to show large numbers such as 500, 1000, 2000 and above on graph, then he has to consider 1 cm long line on graph as 500 unit bar line. Now 1 cm is divided 5 times to represent 500 unit bar line. 1 cm is denoted as 5; 2 cm as 10; 3 cm as 15 and so on. For number 750 a point in between 5 and 10 is mentioned. The same method can be adopted to represent any number.

Grouped data can be represented graphically in any one of the following ways

(1) Histogram (2) Frequency polygon

(3) Frequency curve (4) Relative frequency map

(5) Cumulative frequency curve or ogive and (6) Scatter or dot diagram

1. Histogram. This graph is used for continuous frequency distribution. The width of the class interval marked along with the X-axis, or abscissa. On these length, rectangles of areas proportional to the frequencies of the respective class intervals are erected.

If the class intervals are of equal lengths, then the heights of the rectangles are proportional to the corresponding frequencies and for unequal class intervals, the heights of the rectangles are proportional to the ratios of the frequencies to the width of the corresponding class.

Following grouped data is obtained in an observation of "rate of reproduction" of 50 fishes of a species. Make a Histogram, Frequency polygon and Frequency curve with the help of data provided.

Class intervals	0–10	10–20	20–30	30–40	40–50	50–60	60–70	70–80	80–90
Frequency	3	4	7	8	9	9	2	6	2

Solution OX-axis 1 cm = 10 class interval denoting the rate of reproduction

OY-axis 1 cm = 1 frequency representing the frequency of rate of reproduction

The frequency of 1st class interval 0 – 10 is 3 which is being represented by 3 cm. = 30 small squares on OY axis because 10 small squares = 1 frequency. In the same fashion rectangle for each class interval and frequency is plotted and finally a histogram of the above frequency distribution is shown in Fig. 4.1.

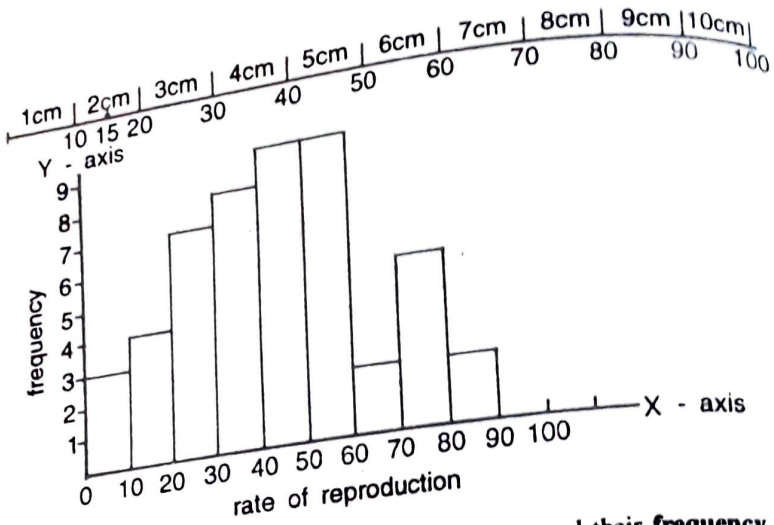


Fig. 4.1. Histogram showing **rate of reproduction** and their **frequency** of 50 fishes of a species.

2. Frequency polygon and 3. frequency curve. The values of the variable for an ungrouped data are taken as the abscissae and their frequencies are taken as the ordinates. For a grouped data, the mid-points of the class intervals are taken as the abscissae. Then a *frequency polygon* is obtained by joining the plotted points by the straight lines. If the class intervals are of small length, then the plotted points are joined by free hand. The curve so obatined is known

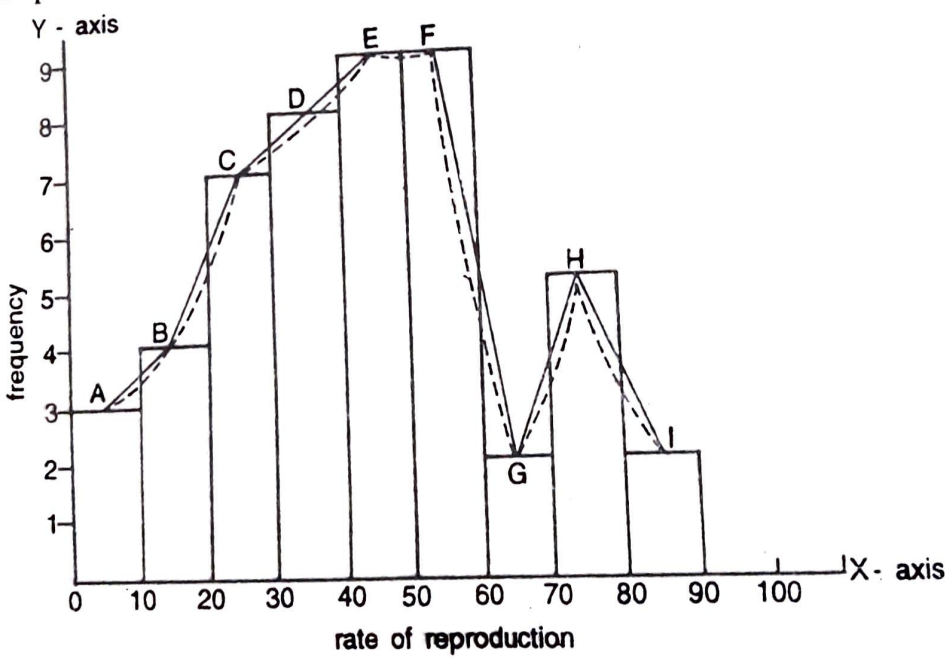


Fig. 4.2. Frequency polygon and frequency curve showing **rate of reproduction** and their **Frequency** of 50 fishes of a species of fish.

Unbroken lines joining mid points A, B, C, D, E, F, G, H and I of rectangle show the **frequency polygon**

Broken lines joining mid points A to I of rectangle represent the **frequency curve.** It is drawn by joining the mid-points of class intervals of upper horizontal lines of rectangle by free hand.

at a glance. Numerous diagrams are used in biometric analysis. Important types of diagrams used for presentation of data are given below:

(I) Line diagrams

(II) Bar diagrams

(III) Pie-diagrams or Pie chart

[I] Line diagrams

This is the simplest type of diagram. For diagrammatic representation of data, the frequencies of the discrete variable can be presented by a line diagram. The variable is taken on the X-axis, and the frequencies of the observation on the Y-axis. The straight lines are drawn whose lengths are proportional to the frequencies.

Worked example **:** The frequency distribution of a discrete variable (Rat of reproduction of 50 fishes) is given in the following table 4.14.

Table 4.14

Rate of reproduction	10	20	30	40	50	60	70	80	90
Frequency	3	4	7	8	9	9	2	6	2

The line diagram is given in Fig. 4.7 of the data presented in above Table 4.

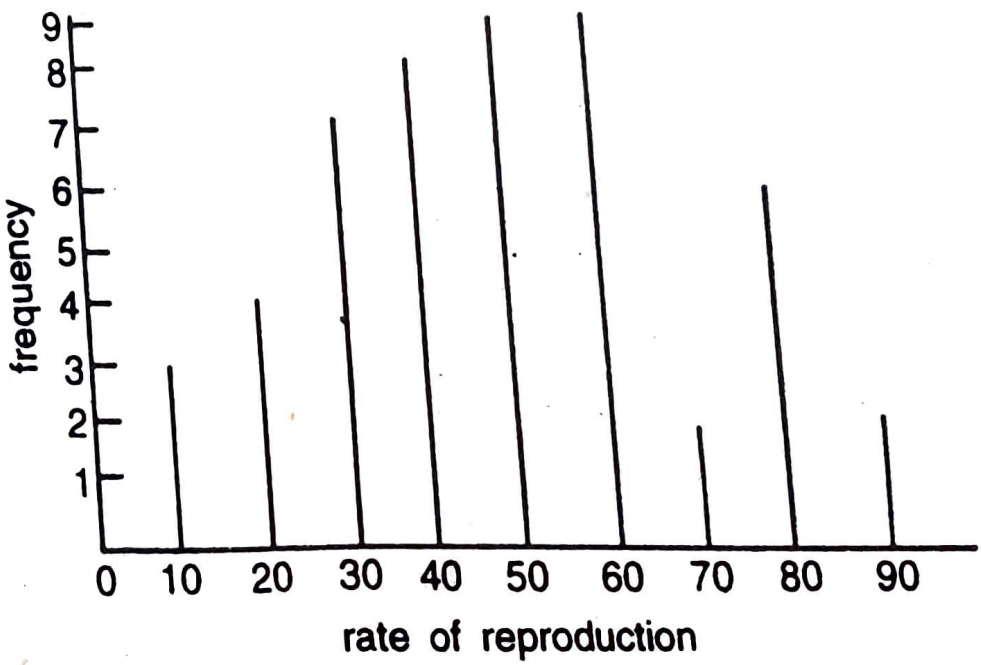


Fig. 4.7. Line diagram.

[II] Bar diagram

Bar diagrams are one dimensional diagrams because the length of the important, and not the width. In this case the rectangular bars of equa is drawn.

diagrams (3) Percentage ... diagrams (4) Divided bar diagrams

1. Simple bar diagram. A simple bar diagram is used to represent one variable. As one bar represents only one figure, there are as many bars the number of figures. For example a simple bar diagram (Fig. 4.8) is drawn taking data of following example.

Worked example. Oxygen consumption in cc/kg/h in different months of year in a species of fish was obtained as below. Draw a simple bar diagram.

Months	J-98	F-98	M-98	A-98	M-98	J-98	J-98	A-98	S-98	O-98	N-98	D-98	J-99
VO_2 consumed	67	74	84	85	100	105	95	90	90	78	74	64	67

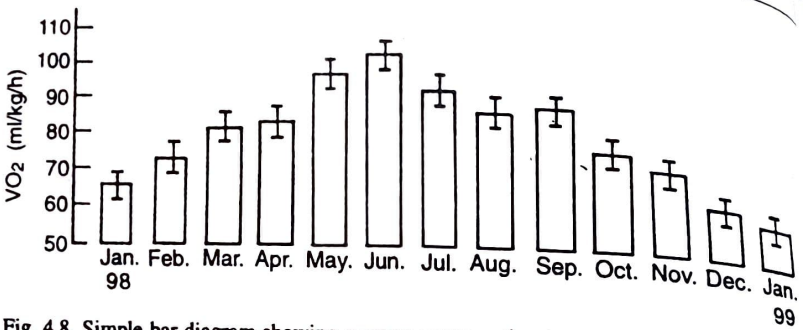


Fig. 4.8. Simple bar diagram showing oxygen consumption in a species of fish recorded in different months of a year.

2. Divided bar diagram. The frequency is divided into different components and such a diagrammatic representation is called a divided bar diagram. Suppose we have to show the average production of four species of fishes in different years, the data can be represented by divided bar diagram. Each bar then would be divided into four parts and each part would represent the mean production of each fish species.

Average catch in metric tonnes of *Wallago, Catla, Cirhinna* & *Clarius* for the year 1993-94, 1994-95, 1995-96 and 1996-97 in India was as follows (Hypothetical data).

Years	Wallago	Catla	Cirhinna	Clarius	Total
1993-94	1383	634	513	400	2930
1994-95	2021	1383	521	313	4238
1995-96	1914	1413	551	900	4578
1996-97	2664	1636	424	265	4989

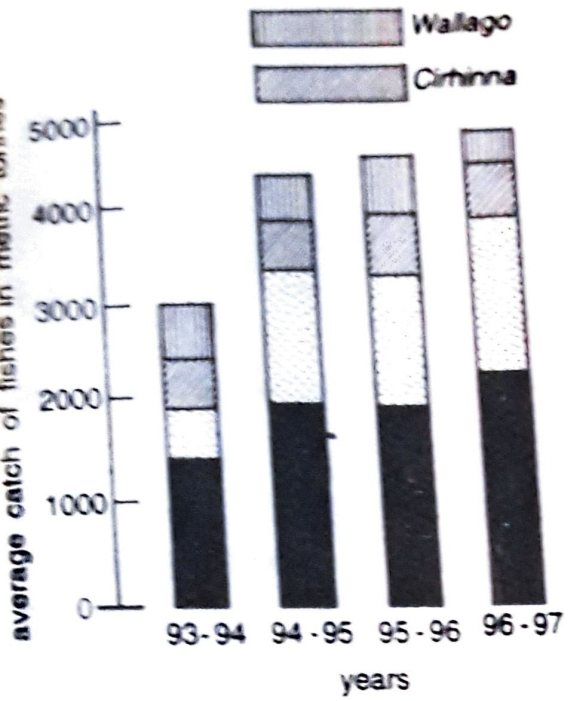


Fig. 4.9. Divided bar diagram representing the production of fishes of four species in four different years.

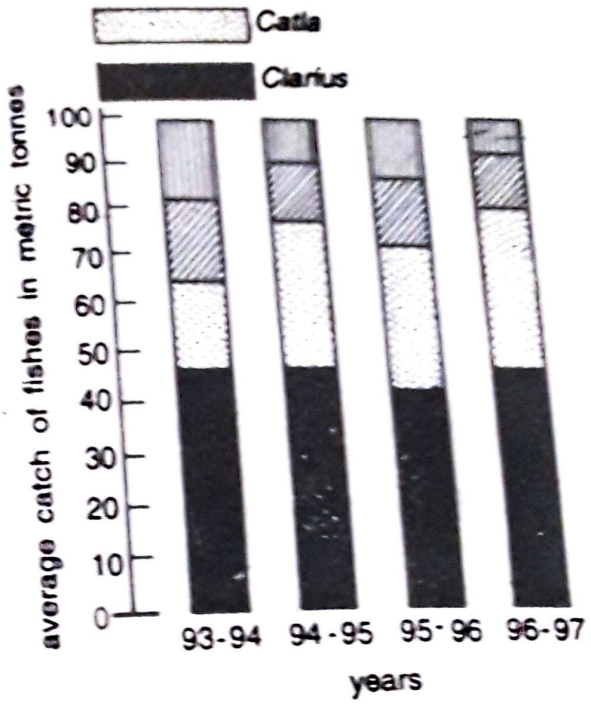


Fig. 4.10. Percentage divided bar diagram representing the data of fish production of 4 species in 4 different years.

3. Percentage **bar** diagram. The length of bars is kept equal to 100 and the divisions of the bar correspond to the percentage of different components. Each component of the bar diagram indicates the average catch of fishes. Above percentage divided bar diagram (Fig. 4.10) is drawn to represent the above data.

4. Multiple bar diagram. When a comparison between two or more related variables has to be made, then multiple bar diagrams are preferred. The technique of simple bar diagrams can be extended to represent two or more sets of interrelated data in a diagram.

Worked example : Value of three haematological parameters viz., RBCs count, Hb% and PCV of a species of fish was studied for 13 months (Between Jan '95 to Jan '96). Data obtained is given below to draw a multiple bar diagram.

Months	J-95	F-95	M-95	A-95	M-95	J-95	J-95	A-95	S-95	O-95	N-95	D-95	J-96
RBCs (Lac/mm^3)	2.01	2.01	2.08	2.12	2.25	2.46	2.27	1.87	1.91	2.30	2.19	2.12	2.04
Hb% (mg/100ml)	8.5	8.6	8.8	9.1	11.7	12.6	11.8	9.7	9.6	12.2	11.8	10.9	8.6
PCV (%)	14.1	14.1	14.1	14.4	16.6	19.6	26.2	24.9	14.4	14.5	25.6	24.2	14.0

[III] Pie chart (π chart) or pie diagram or sector diagram

It is an easy way of presenting discrete data of qualitative characters such as blood groups, Rh factors, age groups, sex group etc. The frequencies of the groups are shown in a circle. Degrees of angle denote the frequency and area of the sector. It presents comparative difference at a glance. Size of each angle is calculated by multiplying the class percentage with 3.6 i.e., 360/100 or by the following formula :

$$\text{Size of the angle} = \frac{\text{Class frequency}}{\text{Total observations}} \times 360^{\circ}$$

(Pie chart always represents the data in percentage)

Worked example. In a study of blood groups in 1629 males and 1181 females of Bihar state following data were obtained —

Table 4.16.

Blood groups	No. of persons			Percentage	Degrees
	Male	Female	Total		
A	427	317	744	26.5	94.4
B	559	412	971.	34.5	124.2
O	521	367	888	31.6	113.8
AB	122	85	207	7.4	26.6
Total	1629	1181	2810	100.0	360.0

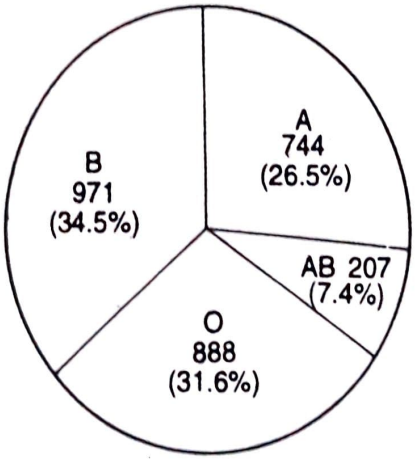


Fig. 4.13. Pie chart or sector diagram showing distribution of blood groups as given in table 4.16.

Size of angle for blood group A in table $= 26.5 \times 3.6$

$$= 95.4 \text{ or } \frac{744}{2810} \times 360 = 95.4$$

Pie chart can be drawn showing distribution of blood groups as given table 4.16.

numbe... and so on.

Sample. The selected part of a population is known as sample. In statistics, population always means the total number of individual observations from which inferences are to be made at a particular time. A sample represents the small collection of the population which has actually been observed. For example, all patients of AIDS of the world represents a population, whereas individual observations on 10 or 20 or 30 (any convenient number) patients from the population refer to a sample, as detailed in chapter 3.

Data. A set of values recorded for an event is called data. The data in statistics are generally based on individual observations. The Hb% of 10 patients suffering from Kalazar was measured as 10.2, 9.6, 8.8, 10.7, 9.9, 10.8, 11.3, 9.5, 8.9, 8.8, mg/100 ml. Here 10.2, 9.6.....8.8 mg/100 ml. are a set of values for an event i.e. Hb% and is called data. While studying statistical methods, we come across numerous data. The data collected for statistical analysis do not consist of observations that are identical, since there would be little reason to study such a variation. The data counted for the purpose of analysis will represent the varying values of variable i.e. a characteristic that shows variation.

A data collected by personal investigation from the original source, by performing some experiments is called **primary data**. For biological researches, data collected only from personal experimental study i.e. primary data is used. Data is of two types —

(a) ***Qualitative.*** According to quality or attributes the data is called qualitative. For example, lions of **Gir sanctuary** of Gujrat State are to be classified in respect to one attribute say sex, in two groups, one is of **male** and the other is of **female**.

(b) Quantitative. According to magnitude the data is called quantitative For example, chickens of a poultry farm may be classified on the basis of thei growth rate. Quantitative data may also be classified into two types —

(i) Continuous. Values of variate do not exhibit any breaks or jumps. F example the increasing length and weight of a child.

(ii) Discrete. Values of variate vary by infinite jumps. For example oxygen consumption of rat (*Rattus rattus*) of different weight groups measured as 500 cc/h/100 ml, 600cc/h/100 ml, 620 cc/h/100 ml, 680 cc/h/ ml and so on.

Observation. Measurement of an event is called observation. For insta blood pressure, temperature of body, oxygen consumption etc. are e whereas, 160 mm & 80 mm (upper and lower pressure), 106°F, 65 kg/hou ml are their respective observations. The source that gives observations su

is not possible to observe all the values of a variable, it is infinite. An infinite population is unlimited in size e.g. the number of RBC's in human body, the number of zooplanktons in sea, and so on.

Sample. The selected part of a population is known as sample. In statistics, population always means the total number of individual observations from which inferences are to be made at a particular time. A sample represents the small collection of the population which has actually been observed. For example, all patients of AIDS of the world represents a population, whereas individual observations on 10 or 20 or 30 (any convenient number) patients from the population refer to a sample, as detailed in chapter 3.

Data. A set of values recorded for an event is called data. The data in statistics are generally based on individual observations. The Hb% of 10 patients suffering from Kalazar was measured as 10.2, 9.6, 8.8, 10.7, 9.9, 10.8, 11.3, 9.5, 8.9, 8.8, mg/100 ml. Here 10.2, 9.6.....8.8 mg/100 ml. are a set of values for an event i.e. Hb% and is called data. While studying statistical methods, we come across numerous data. The data collected for statistical analysis do not consist of observations that are identical, since there would be little reason to study such a variation. The data counted for the purpose of analysis will represent the varying values of variable i.e. a characteristic that shows variation.

A data collected by personal investigation from the original source, by performing some experiments is called **primary data**. For biological researches, data collected only from personal experimental study i.e. primary data is used. Data is of two types —

(a) ***Qualitative.*** According to quality or attributes the data is called qualitative. For example, lions of **Gir sanctuary** of Gujrat State are to be classified in respect to one attribute say sex, in two groups, one is of **male** and the other is of **female**.

(b) ***Quantitative.*** According to magnitude the data is called quantitative. For example, chickens of a poultry farm may be classified on the basis of their growth rate. Quantitative data may also be classified into two types —

(i) Continuous. Values of variate do not exhibit any breaks or jumps. For example the increasing length and weight of a child.

(ii) Discrete. Values of variate vary by infinite jumps. For example oxygen consumption of rat (*Rattus rattus*) of different weight groups measured as 500 cc/h/100 ml, 600cc/h/100 ml, 620 cc/h/100 ml, 680 cc/h/100 ml and so on.

Observation. Measurement of an event is called observation. For instance blood pressure, temperature of body, oxygen consumption etc. are observations whereas, 160 mm & 80 mm (upper and lower pressure), 106°F, 65 kg/hour, 500 cc/h/100 ml are their respective observations. The source that gives observations