2.2 Collection and presentation of data

DATA:

Data is a collection of observations expressed in numerical figures. The collection may be done

in *two* ways.

(*a*) by complete enumeration and

(*b*) simple survey method.

Data is always in collective sense and never be used singular.

Types of Data:

The statistical data can be divided into *two* broad categories:

(*a*) Qualitative (*b*) Quantitative.

Qualitative Data :

In this type of data, there is no numerical relation with one another.

Example: Skin colour—brown, black, white

Eye colour—blue, brown

Sex—Male, Female.

Quantitative Data:

1. In this type of data, there is numerical relation with one another.

2. It may be continuous or discrete.

Example: Discrete = Number of books, number of students.

Continuous = Height or Weight of person.

Qualitative data	*Quantitative data*
1. Always Discrete.	. 1. Discrete or continuous.
2. No magnitude	2. Have magnitude
3. Persons with same character are counted to form groups	3. Arranged by both character and frequency.
4. Results are expressed as ratio or proportion.	4. Such data are analysed through statistical method *e.g.*, mean, median, mode, S.D. etc.

A. According to source of data collection:

(*a*) **Primary data:** Directly from field or experiment.

(*b*) **Secondary data:** Obtained from primary data or review.

B. According to variable:

(*a*) Univariable.

(*b*) Bivariable.

(*c*) Multivariable.

C. According to compilation:

(*a*) **Raw data:** Data before compilation.

(*b*) **Derived data:** Calculated from primary value of data.

PRIMARY DATA:

These data are collected directly from the field of enquiry for a specific purpose. These are raw data or data in original nature, and directly collected from population. The collection of primary data may be made through either by complete enumeration or sampling survey methods.

SECONDARY DATA:

These are numerical information which have been already collected by some agency for a specific purpose and are subsequently compiled from that source for application in different connection. In other words, data used by any other agency than the collecting authority will be termed as secondary data.

COLLECTION OF PRIMARY DATA:

The following methods are generally used for collection of primary data:

(*a*) Direct personal observation.

(*b*) Indirect oral investigation.

(*c*) Questionnaires sent by mail.

(*d*) Schedules sent through investigators.

QUESTIONNAIRE:

It is a proforma containing a sequence of questions relevant to a statistical enquiry. It is used for collection of primary data from individual persons through their response to the set of questions.

RELATIVE ADVANTAGES OF PRIMARY DATA:

1. Primary data provides with detailed information but in secondary data some information may be suppressed.

2. Primary data is free from transcribing errors and estimation errors where as a secondary may contain such errors.

3. Secondary data normally do not contain information regarding methods of procuring data where as primary data often include them.

4. Cost effectiveness is a vital plus point for using secondary data. Thus time, cost suitability and accuracy are the essential factors whether we would use primary or secondary data.

POPULATION:

It is an entire group of people or study elements-persons, things or measurements having some common fundamental characteristics.

(*a*) **Finite:** If a population consist of fixed number of value, it is said to be finite *e.g.*, number

of days in a week.

(*b*) **Infinite:** If a population consist of an endless succession of values, it is said to be infinite

e.g., number of animals in ocean.

SAMPLING:

Target population → Study population → Study sample
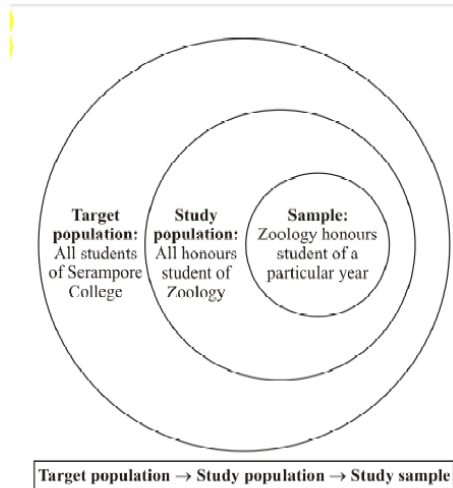
The technique of obtaining information about the whole group by examining only the part of the whole group is called *sampling*.

Types of Sampling:

(*a*) Random sample (Probability sample).

(*b*) Non-random sample (Non-probabilities sample).

Objectives of Sampling:

1. Estimation of population parameter (mean,SD etc.) from the sample statistics.

2. To test hypothesis about the population from which the sample or samples are drawn.

SAMPLE:

It is a relatively small group of selected number of individuals or objects or cases drawn from a particular population and is used to throw light on the population characteristics.

RANDOM SAMPLE:

It is a sample chosen in a very specific way and has been selected in such a way that every element in the population has an equal opportunity (unbiased) of being included in the sample.

CHARACTERISTIC:

The term 'characteristic' means a quality possessed by an individual (*i.e.*, object, item of population). Height, weight, age etc. are characteristics.

In statistics, characteristics are of *two* kinds:

(*a*) Non-measurable 'characteristics' (attributes)

(*b*) Measurable 'characteristics' (variables).

(*a*) **Attributes:** Attributes are the non-measurable characteristics which can not be numerically

expressed in terms of unit. These are qualitative object.

For example—Religion, Nationality, Illiteracy etc.

(*b*) **Variables:** Variables are the measurable characteristics which can be numerically expressed in terms of some unit. These are quantities which are capable of being measured by quantitative methods directly.

For example—Height in inches, cm, weight in kg, pound, marks in examination etc.

(*i*) **Discrete Variables (Discontinuous/meristic):** There are the quantities which can be measured in whole integral values. It does not take fractional value.

Example—Number of books marks in examination.

(*ii*) **Continuous Variables:** These are quantities which can take any value in specified range. Thus it can take integral and fractional values both.

Example—Heights, weights etc.

Discrete Variables:

No .of the mango tree	No of Mango to the tree
1	30
2	70
3	100
4	80
5	60

Continuous Variable:

Specimens (Mango fruit)	Weight in Grams
1	50.26
2	37.32
3	48.75
4	37.81
5	60.23
6	54.55

STATISTICAL ERROR:

In statistical terminology the word 'error' is used in special sense. Error shows the extent to which the observed value of a quantity exceeds the true value.

Error = Observed value – True value.

TYPES:

In statistical terminology the word 'error' is used in special sense. Error shows the extent to which the observed value of a quantity exceeds the true value.

Error = Observed value – True value.

TYPES:

Statistical error may be classified as:

(*a*) *Biased errors:* (which arise due to personal prejudices or bias of investigator & informants).

(*b*) *Unbiased errors:* (which enter into statistical enquiry due to chance causes).

ARRAY:

The presentation of data in ascending order of magnitude is called *array*.

TALLY:

(*i*) A tally mark is an upword slanted stroke (*I*) which is put against each occurrence of value.

(*ii*) When value occurs more than four times the fifth occurrence is denoted by a cross (\) tally mark, running diagonally, across the four tally marks. This facilitates the counting of tally marks at the end.

(*iii*) The total count of tally against each value is called its *frequency*.

(*iv*) A frequency distribution with individual values is called *simple frequency distribution*.

Example: Form a frequency table for the following variables:

51, 59, 52, 51, 60, 68, 63, 64, 65, 66, 68, 52, 59 60, 58, 51,

54, 55, 56, 61, 62, 69, 70, 58, 69, 65 67, 63, 63, 62, 61, 51,

59, 63, 68, 67, 69, 53, 53 51, 59, 56, 55, 70, 65, 62, 65, 66,

69, 70, 52, 55 64, 65, 69, 61, 63, 54, 64, 61, 61, 62, 51, 52,

52, 54, 55, 52, 52, 66.

Solution:

Values of variables	Tally	Frequency
51								6
52									7
53				2
54					3
55						4
56				2
58				2
59						4
60				2
61							5
62						4
63							5
64					3
65							5
66					3
67				2
68					3
69							5
70					3
Total		70

CLASSIFICATION: It is the process of arranging the collected statistical information under different categories or classes according to some common characteristics possessed by an individual member.

Types of Classification:

There are four types of classification.

(*a*) **On qualitative basis:** Here non measurable characteristics are classified.

(*b*) **On quantitative basis:** Here measurable characteristics are classified.

(*c*) **On time basis:** Here the statistical data are arranged in order of their time of occurrence.

(*d*) **On the geographical basis:** The total population of a country may be classified by states, or districts. The basis of classification in such cases is by geographical regions.

METHOD OF PRESENTATION OF STATISTICAL DATA:

Statistical data are presented in *three* processes:

(*a*) **Textual Presentation:**

(*i*) Numerical data presented in a descriptive form are called *textual presentation*.

(*ii*) It is lengthy. Some words may repeat several times in the text.

(*iii*) It becomes difficult to grasp salient points in a textual presentation.

(*b*) **Tabular Presentation:**

(*i*) The logical and systematic presentation of numerical data in rows and columns designed to simplify the presentation and facilitate comparison is termed as tabulation.

(*ii*) Tabulation is thus a form of presenting quantitative data in condensed and coincise form so that numerical figures are capable of easy & quick reception by the eyes.

(*iii*) It is more convenient than textual presentation.

(*c*) **Graphical Presentation:** The presentation of quantitative data by graphs and charts are termed as graphical presentation.

Tabulation:

It may be defined as the logical and systematic presentation of numerical data in rows and columns designed to simplify the presentation and facilitate comparisons.

The Advantages of tabulation are:

(*i*) It enables the significance of data readily understood and leaves a lasting impression than textual impression.

(*ii*) It facilitates quick comparison of statistical data shown between rows and columns.

(*iii*) Errors and omissions can be readily detected when data are tabulated.

(*iv*) Repetition of explanatory terms and phrases can be avoided, and the concise tabular form clearly reveals the characteristics of data.

Types of Tabulation:

There are *two* types of tabulation:

(*a*) **Simple tabulation:** It contains data in respect of one characteristic only.

(*b*) **Complex tabulation:** It contains data of more than one characteristics.

Example: (Simple tabulation): Number of students in three colleges.
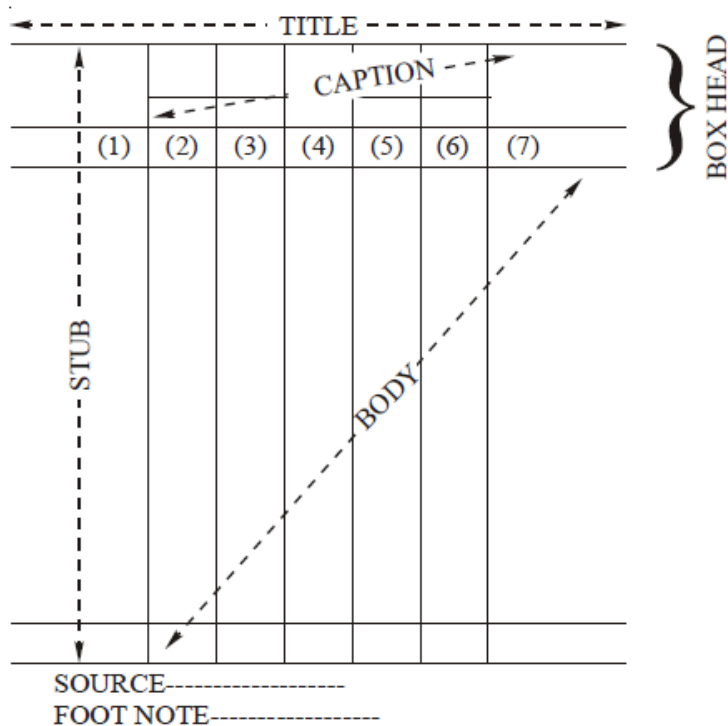
Fig. 1.2. Different parts of table.

Statistical Tables:

Statistical table is a systematic arrangement of quantitative data under appropriate heads in rows and columns. After the data have been collected, they should be tabulated that is put in the form of a table, so that whole information can be had at a glance.

Parts of a Table:

(I) Title:

(*a*) This is a brief description of the contents of the table along with time, place and category of item if required.

(*b*) The title should be clear and precise.

(*c*) It should be at the top of the table.

(II) Stub:

(*a*) The extreme left part of the table where descriptions of the rows are shown in called *stub*.

(*b*) It must be precise and clear.

(III) Caption and Box head:

(*a*) The upper part of the table which shows the description of columns and sub columns is called *Caption*.

(*b*) The whole of the upper part including caption units of measurement and column number if any is called *Boxhead*.

(IV) Body:

(*a*) It is the main part of the table except the title stub and captions.

(*b*) It contains numerical information which are arranged in the table according to the descriptions of the rows and columns given the stub and caption.

(V) Source and foot note:

(*a*) It is customary that source of data from which information has been arrived should be given at the end of the table.

(*b*) Foot note is the part below the body where the source of data and any explanation are shown.

Essential features of a good table:

1. A table must have a title giving clear and precise idea about the contents of the table.

2. Units of measurements adopted in a table must be shown clearly in the top of the column.

3. It is a necessity that an investigator prepares a table well proportioned in length and breadth.

4. For a compatible comparison, column of relevant figures must be kept as close as possible.

5. Distinction is preferred in columns and sub columns. It can be made by distinct ruling (viz double ruling, single ruling etc.).

6. Totals of columns may be shown in the bottom of the table. In cases where row totals are useful, they should also be shown.

7. Table must contain necessary details.

8. Source of information must be disclosed at the end of the table.

9. Any ambiguous or confusing entry in the table should bear a special note at the end of the table for experiment.

10. The arrangement of items in the table should have a logical sequence.