COLLECTION OF DATA.

A statistical enquiry of a phenomenon always begins with collection of data. 1 of the imp () of a statistician is to arrange & collect most reliable & accurate data related to the topic of study.

Data may be obtained either from primary source / secondary source.

The source of data are said to be primary when data are published / released by the organisations that actually collected them.

The source of data are secondary, when data are published by anthr organisation.

Depending on the source statistical data are classified under 2 catagories → primary data & secondary (d)

* The primery (d) are those which are collected for the 1st time & are thus original in character.
* Sec. (d) are those which hve already been collected by someother persons & which hve passed through the statistical machine atleast once.

=> Methods of collecting p. data :-

1) Direct personal Investigation :-

There is a face to face contact with the persons from whome the info- is to be obtained.

* Merits :-
- original data are collected.
- Reliable info.
- More accurate
- Uniformity in the collection of info.

* Demerits :-
- It can be used only when the area of investigation is small.
- More expensive & time consuming.

2) Indirect oral investigation :-

* we have to collect data from those persons who may possess some knowledge about the enquiry.
* This method is suitable when the area of investigation is very large.

-> Merits :-
* Large area can be covered.
* More complex data can be analysed.
* Simple & convinient.
* saves money, time & labour.

→ Demerits :-

* It may not always be accurate & ~~incorrect~~ true.

3) Schedules sent through enumerators :-

In this method of collecting data, is that of sending schedules through enumerators.

The (enu) contact the informence, get replace to the questions conveyed in a schedule & fill them in their own hand writing in a questionies form.

4) Mailed questionies Method :-

The (ques) mailed to the informance. It is also carries a request for the quick responds & ~~this~~ lays down the aims & objectives of collecting the info.

5) Information through correspondent :-

In this method the investigator does not collect the info directly frm the respondence.

The task is handled to a few persons/ grp of persons → correspondents.

→ Drafting of the questionnaire :-

* distinction is ~~ofte~~ made b/w (q) & shedule. The (q) is filled up by the informant, while shedule is filled by a trained enumerator who gathers the info by questioning informnts.

The follo...are the points to be borne in mind while drafting a gd (q)/shedule-

1) should be as few questions as possible.
2) (q) should be simple & easy to understnd.
3) (q) should not be ambiguous.
4) (q) should follow a logical sequence.
5) (q) " be such that they will be answered without bias.
6) (q) " not be unnecessarily inquisitorial
8) specific info (q) should be included.
9) open (q) ought to be avoided.
10) Instructions to be informants should be given.
11) (q) should be capable of objective ans
12) (q)r should ~~be~~ attractive
13) pre testing the (q)r [questionaire]
14) (q) requiring calculations should not be asked.

→ Secondary data :-

* official publications of central, state & local governments.

2) Reports of committees & commissions on various enquiries.
3) Reports & publications of trade associations, banks, chamber of commerce,
5) official publications of govnmnt bodies (RBI)
6) " " of international bodies like WHO, IMF.

Since S. data have already been collected it is highly desirable that a proper scrutiny of such data is made bfre they are used by investigtr.

1] S. data must be reliable :-
Investigatr should see to it that the data has not been affected by bias of original investigatr.

2] S. data should be suitable :-
(I) → (investigatr)
when gathering S. data the (I) must give particular attention to the (q).

3] S. data should be adequate :-
S. data may be reliable & suitable bt it may not be adequate for the purpose of enquiry.

Caution—

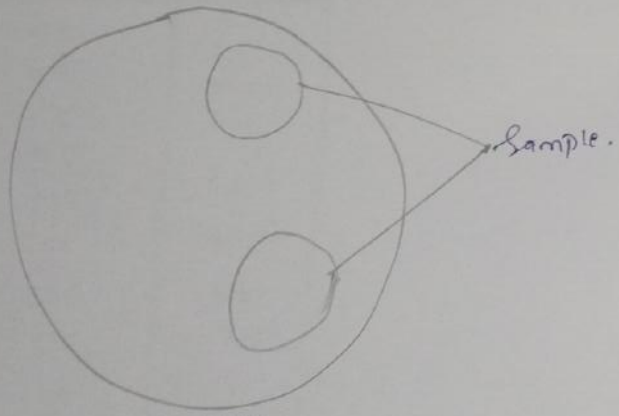
→ Concepts of population & sample :-

* Population :- (P).

It is the entire grp that you want to draw conclusion about.

* Sample :-

It is the specific grp that you will collect data from.
Size of the sample is always < than the total size of the (P).



Population

Sample is the small part of P.

→ Census & Sample Method :-

* If we are studying each & every element of the (P) the process → census method
* If we are studying only a sample the process → Sample survey, sample

Method / Sampling.

→ Variables & attributes :-

* Quantitative / Numerical characteristics of data → Variables
* A variable (V) is a quantity which varies from (individual observation to another with in a given set of values → its domain.
eg = height in cm, weight in kg...
* A (V) whose domain retains only 1 value in a particular situation → Constant
* The particular numerical value of a variable from a given domain is real no. → variate.
* A (V) is said to be discrete when its value can assume only integral no. & it cannot be measured but counted only.
* A (V) which can assume any value within a given domain → contin (V).
* Data which are of qualitative characteristics → Attributes
* Attributes cannot have direct numerical measurement.
eg = Sex, religion, nationality, literacy..., are the attributes of a person which cannot be measured numerically.

→ Types of data :-

* Data can be classified according to variables & attributes. → (A)
classification according to (v) is → quantitative classification & corresponding data → (q) data.
It is also called numerical data.

* In (A)-ive classi.. the data are dassi.. on the basis of quality. → qualitative data

1) Time series data = A T.S a set of obsr^v. on the values that a variable takes at diff.. times.

2) Cross section data = They are data on 1 / more (v). collected on same point in time

3) pooled data = It is the combination of both time series & c.section data.

4) Nominal & ordinal data :- For a eg we can record sex of a person o & 1 corresponding to a male / female. In this artificial / nominal way categorical data can be made into numerical data. & refer to the non as nominal data.

N. Data are numerical in name only. The assigned no.. cannot be added, subtracted, etc. The only arithmetic operation that can be carried out are the count of each category. In these situation when we cannot do anything except setup in equality we refer to the data as ordinal data.

Eg of ordinal scale include - quality ranking, rankings of a games in a tournament .. .

As the ordinal scale measurement is ↑ than the nominal scal measrmnt, all the statistical techniques which are applicable in the case of N.scale measrmnt can also be used. for the o. scale measrment.