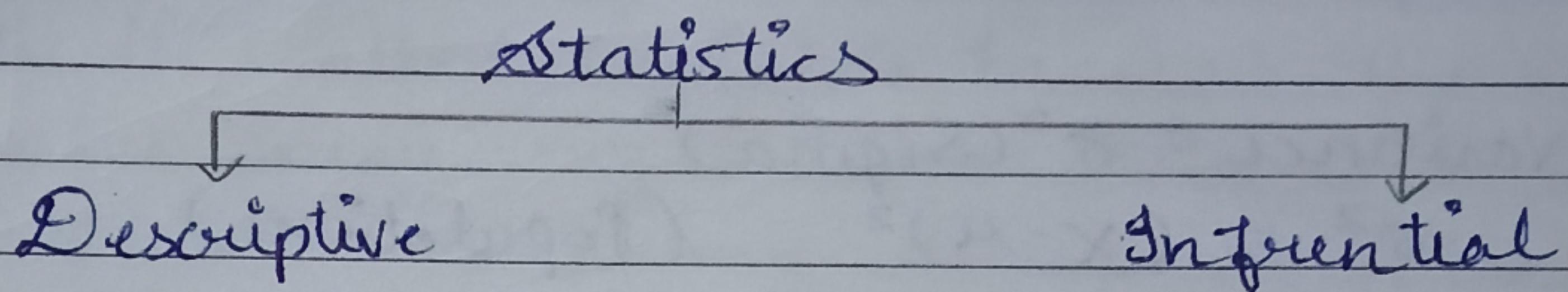


Statistics

Statistics is a branch of science of collecting, organizing and analysing data. There are two types of statistics.



- Descriptive: It consists of organizing and summarizing data in the form of graphs, charts and numbers.

Techniques to measure

1. Measure of central tendency.
→ Mean, Mode, Median

- Ex:-

$$x = 3, 2, 5, 10, 6$$

Mean: μ = Population, \bar{x} = Sample

$$\bar{x} = \frac{\sum x}{n} \quad (\text{Sum of observations})$$

n (Total no. of observations)

$$\bar{x} = \frac{26}{5} = 5.2 \approx 5$$

Median: 2, 3, 5, 6, 10
= 5

For even = $(\frac{n+1}{2})^{\text{th}}$ term

Reg
Ans
Ques

Mode: highest occurring no.

If there are 2 modes then it is called bimodal.

2. Measure of dispersion.

Range, Variance, Standard Deviation

• Variance : σ^2 (Sigma)²

$$\sigma^2 = \frac{(x - \mu)^2}{n-1} \quad (\text{Population})$$

$$s^2 = \frac{(x - \bar{x})^2}{N} \quad (\text{Sample})$$

• $SD = \sqrt{\sigma^2}$

$$\sqrt{\sigma^2} = \sqrt{\frac{(x - \mu)^2}{n-1}} \quad (\text{population})$$

$$\sqrt{s^2} = \sqrt{\frac{(x - \bar{x})^2}{N}} \quad (\text{sample})$$

3. Probability Density function (PDF), Probability Mass function (PMF), Correlation

• Inferential : It consists of using data, you have major to form conclusions, hypothesis, predictions.

Techniques to measure.

1. Z-test

4. Anova test or

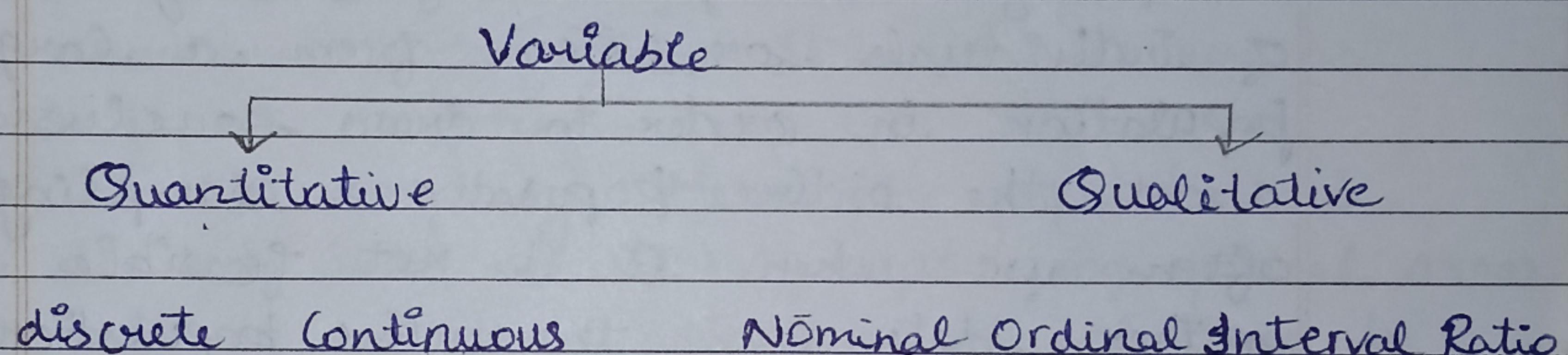
2. t-test

5. F-test

3. Chi-sq test

Variable and types

Variable: A variable is a characteristics, numbers or quantity that can be measured or counted. It is also called data item. It is called variable because the variable value may vary between data units in a population and may change in value over time.
Ex:- age, sex, religion, income etc.



1. **Quantitative:** Quantitative are those variables that can be counted or measured.
Ex :- 121, 15.1.
 - (i) **Discrete** :- 0, 1, 2, 3, 4, 5 without any fraction.
 - (ii) **Continuous** :- 1.1, 1.2, 1.3, 2.2 having a fraction.
2. **Qualitative:** Qualitative are those which we can define a variable on the base of quality. **Ex:-** Sex (M, F), True / False.
 - (i) **Nominal** :- It is in categories (Female, Male). It is a categorical, no ranking and no order. **Ex:-** F, M or M, F.
 - (ii) **Ordinal** :- It is a categorical, nominal, ordered as well as ranking. **Ex:-** Letters (A, B, C, D,)

(iii) Interval : Interval data lacks the absolute zero point. Ex :-

- IQ score :- Meaningful diff b/w 107 and 108 IQ score.
- Temperature :- Meaningful diff b/w 30°C - 40°C

(iv) Ratio : It has a defined zero point.

Ex :- Income, height, weight etc.

Sampling

Sampling is a process of selecting a subset of individuals or items from a large population in order to draw conclusions about the entire populations. Sampling is often used when it is not feasible to gather data from the entire population.

There are two types of sampling

Sampling

Probability sampling
It involves random selection. It is of four types :-

- Simple random sampling.
- Stratified sampling
- Systematic sampling.
- Cluster sampling.

Non-Prob. sampling

It doesn't involve random selection or criteria. It is of four types :-

- Convenience sampling
- Purposive sampling
- Snowball sampling.
- Quota sampling.

(i) Simple random sampling : In this method every member or individual/item in the population has an equal chance of being selected.

- (ii) **Stratified sampling**: It is also called layering sampling. In this population is divided into different sub groups or strata based on certain characteristics. Ex:- income, gender, age etc, then random samples are drawn from each sub groups that represents population.
- (iii) **Systematic sampling**: In systematic sampling every nth element or individual or item is selected from the population after an initial random starting point. It is also known as interval sampling.
- (iv) **Cluster sampling**: Here the population is divided into clusters/groups/geographical areas, and a random selection of clusters is chosen. Then all individuals/items within the selected cluster are included in the sample. This sample is also known as multistage sampling.
- (i) **Convenience sampling**: This method involves selecting individuals or items that are most convenient to access. While it is quick and easy, it may introduce bias (errors) because the sample might not be representative to of the entire population.
- (ii) **Purposive sampling**: It is also called Judge-mental sampling. In this the researcher selects specific individuals/items based on their expertise/purpose about which elements are most relevant to the study.

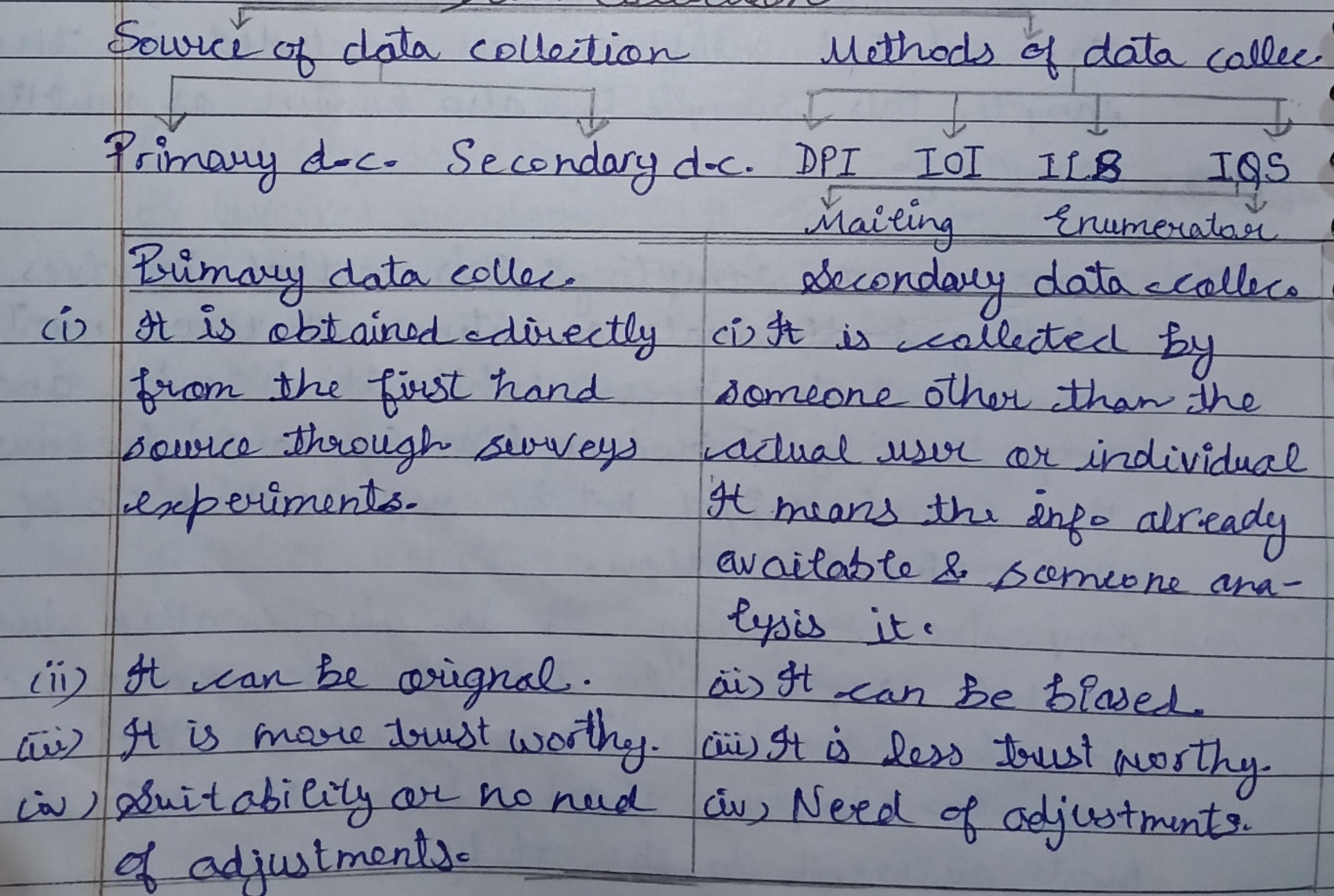
- (iii) Snowball sampling : In snowball sampling the researcher gives some questions to an individual that circulates these questions to other groups or friends.

(iv) Quota sampling : Quota sampling similar to stratified sampling but in quota sampling we select individuals/items on the base of some purpose from the sub groups that represent my population.

Data collection

It is a process of gathering information from different sources in order to find out the solution of a research problem.

Data Collection



- (v) It is costlier. (vi) It is economical.

(i) DPI (Direct Personal Investigation): It is the methods by which data are personally collected by the investigator from the info.

(ii) Indirect oral investigation: It is the method by which info. is obtained not from the persons regarding whom the info is needed - It is collected orally from the other persons who are expected to possess the necessary info.

(iii) Information from local sources or correspondents: Two ways of collecting. Under this method the investigator appoint local persons or correspondents at different places.

(iv) Information through Questionnaires & Schedules: Two ways of collecting info based on questionnaire.

(a) Mailing method: Under this method, questionnaires are mailed to the informants. The method is most suited when → • The area of study is wide.
• The informants are educated.

(b) Enumerator's methods:- Under this method enumerator himself fills the schedules after seeking info. from the informants.

Misuse of Statistics.

- 1 Suspect sample
- 2 Ambiguous/Average
- 3 Implied connection
- 4 Misleading Graphs
- 5 Detached subject
- 6 Changing the subject
- 7 Faulty survey questions.

- (i) Suspect sample :- Data believed to be unreliable or biased. Ex:- Survey responses from non-customers.
- (ii) Ambiguous averages :- Averages that can mislead due to extreme values. Ex:- Averaging income without outliers.
- (iii) Detached statistics :- Data presented without context, leading to misinterpretation. Ex:- Reporting crime rates without considering population growth.
- (iv) Implied connections :- Connections b/w concepts suggested or hinted at. Ex:- Correlation between advertising and sales.
- (v) Misleading graphs :- Graphs that distort or manipulate information. Ex:- Using non-zero baseline on a bar graph.
- (vi) Faulty survey questions :- Poorly designed or leading survey questions. Ex:- Asking "Do you still beat your spouse?"
- (vii) Changing the subject :- Diverting conversation from the current topic. Ex:- Shifting to discussing the weather during work discussion.