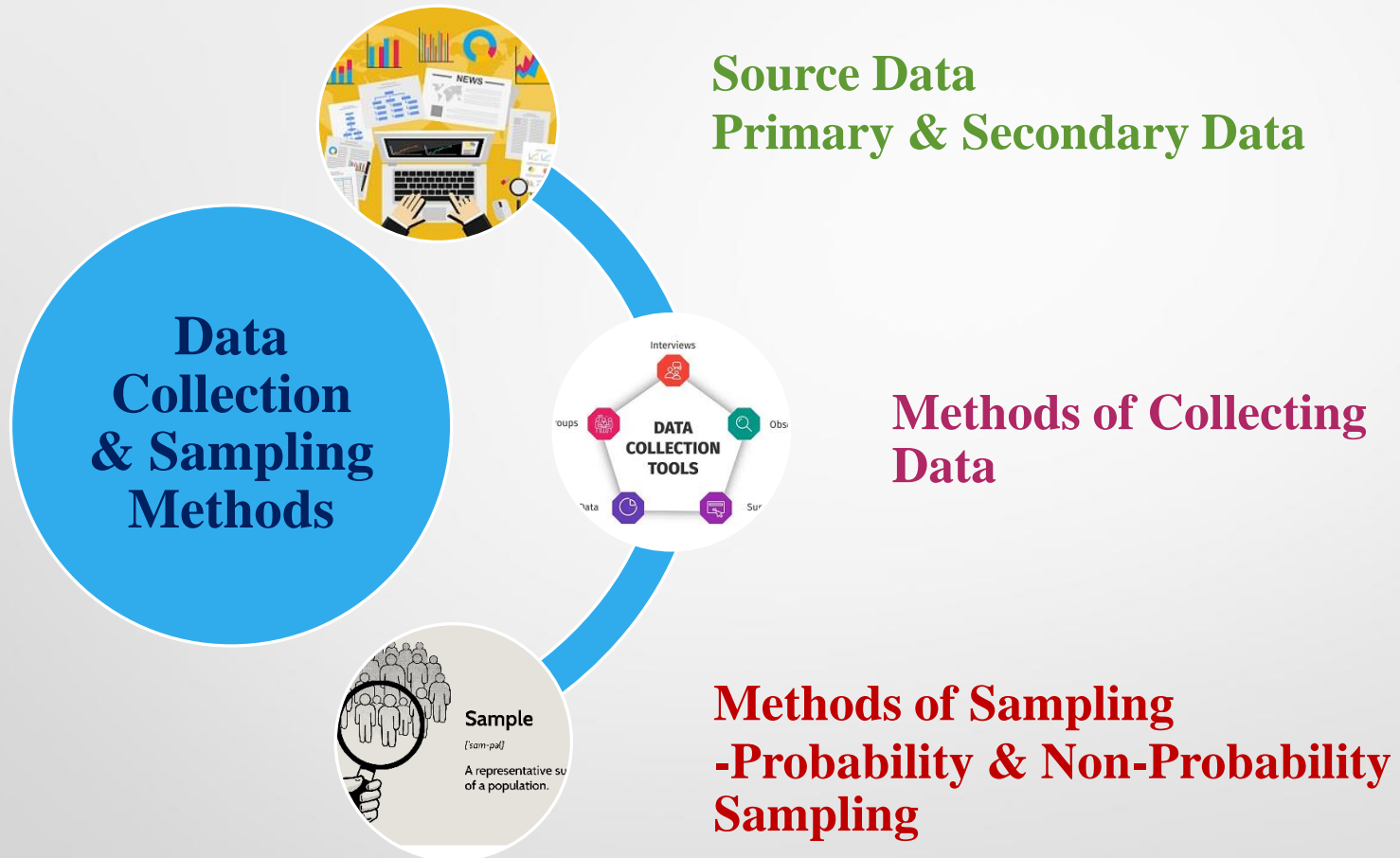


CONTENTS



INTRODUCTION

- **Data collection is a term used to describe a process of preparing and collecting data.**
- **Systematic gathering of data for a particular purpose from various sources, that has been systematically observed, recorded, organized.**
- **Data are the basic inputs to any decision-making process.**

Sources of Data

```
graph TD; A[Sources of Data] --> B[External sources]; A --> C[Internal sources]; B --> D[Primary data]; B --> E[Secondary data];
```

External
sources

Internal
sources

Primary
data

Secondary
data

Two Sources of Data

1. Internal Source:

When data are collected from reports and records of the organization itself, it is known as the internal source.

2. External Source:

When data are collected from outside the organization, it is known as the external source.

Internal & External Sources of Data

Internal sources of Data

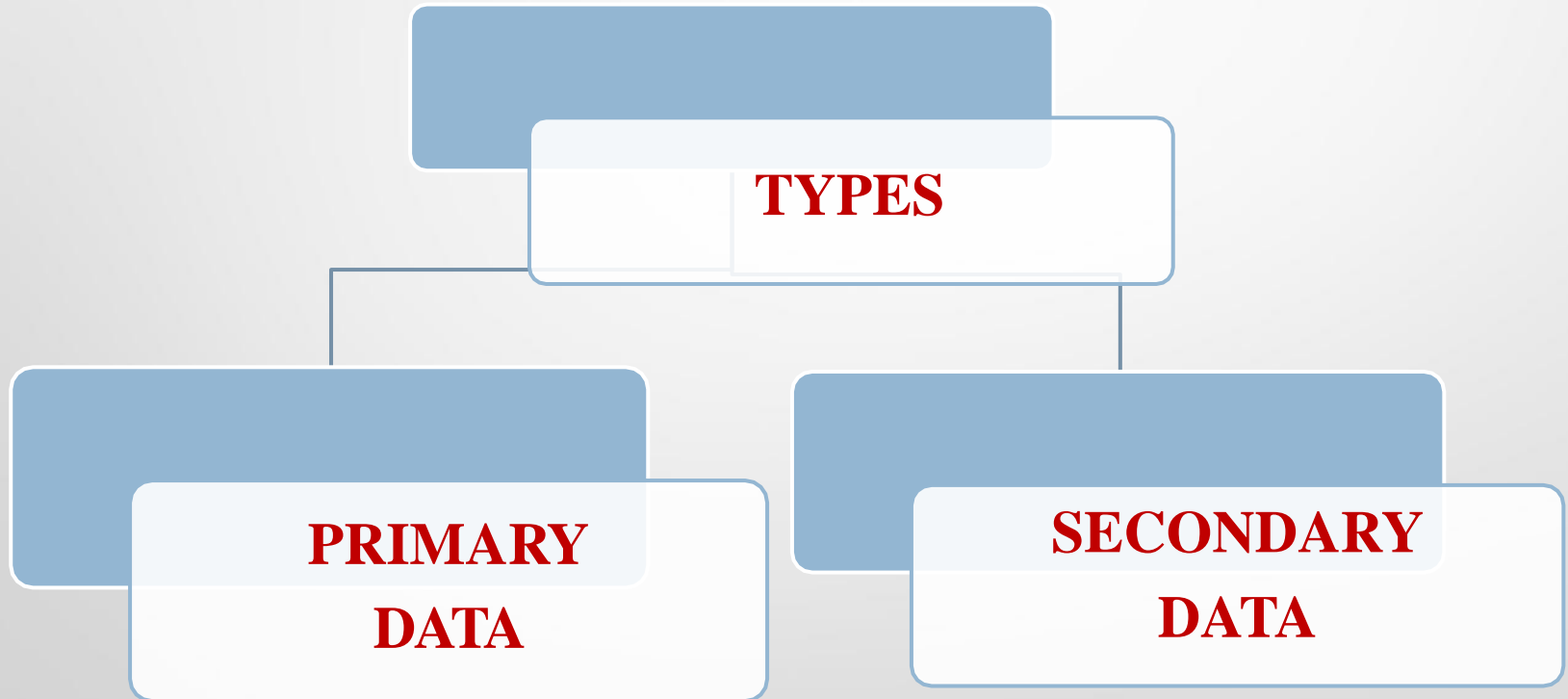
- Many institutions and departments have information about their regular functions, for their own internal purposes.
- When those information are used in any survey is called internal sources of data.
- E.g. social welfare societies.

External sources of data

- When information is collected from outside agencies is called external sources of data.
- Such types of data are either primary or secondary.
- This type of information can be collected by census or sampling method by conducting survey.

Types of Data

Based on method and sources by which the data is collected the data is classified into two types:-



PRIMARY DATA

- The data which are collected from the field under the control and supervision of an investigator.
- Primary data means original data that has been collected specially for the purpose in mind.
- This type of data are generally a fresh and collected for the first time.
- It is useful for current studies as well as for future studies.
- **Example:** your own questionnaire.
- Those data which are collected afresh and for the first time and thus happen to be original in character is known as *Primary data*.

Advantages of Primary Data

- **Original and independent collection increased the authenticity of data.**
- **Directly Collection enhance the reliability of data.**
- **Used in both quantitative and qualitative research methods.**
- **Hidden information can be collected through primary data.**
- **After analyzed primary data can be used as secondary data.**

Disadvantages of Primary Data

- **Reliability depend on respondent's information accuracy.**
- **Information can be bias.**
- **Expensive and time consuming in nature.**
- **Lack of experience among researcher.**
- **Requires field work.**

SECONDARY DATA

- Data gathered and recorded by someone else prior to and for a purpose other than the current project.
- Secondary data is data that is being reused.
- It is the data which may be published or unpublished but has been collected and is used for some other purpose earlier.
- **Example:** data from a book.
- Those data which have been collected by someone else and which have already been passed through the statistical process or analyzed by someone else are known as *Secondary Data*.

Advantages of Secondary Data

- **These data can be quickly manageable.**
- **Time and cost balance remains maintained.**
- **Information available is already analyzed by experts.**
- **Used to update data or reinterpret existing ones.**
- **Helpful for philosopher, thinker or authors for developing new concept.**
- **Field work is less.**

Disadvantages of Secondary Data

- **No standard measurement of validity.**
- **Need expertise.**
- **Accuracy and reliability is always lesser than primary data.**

DIFFERENCE BETWEEN PRIMARY AND SECONDARY DATA

| Bases | Primary Data | Secondary Data |
|--------------------|--|--|
| Originality | Primary data are always original as it is collected by the investigator himself. | Secondary data lacks originality. The investigator makes use of the data collected by other agencies. |
| Suitability | Suitability of the primary data will be positive because it has been systematically collected | Secondary data may or may not suit the objects of enquiry. |



**Time
&
Money**

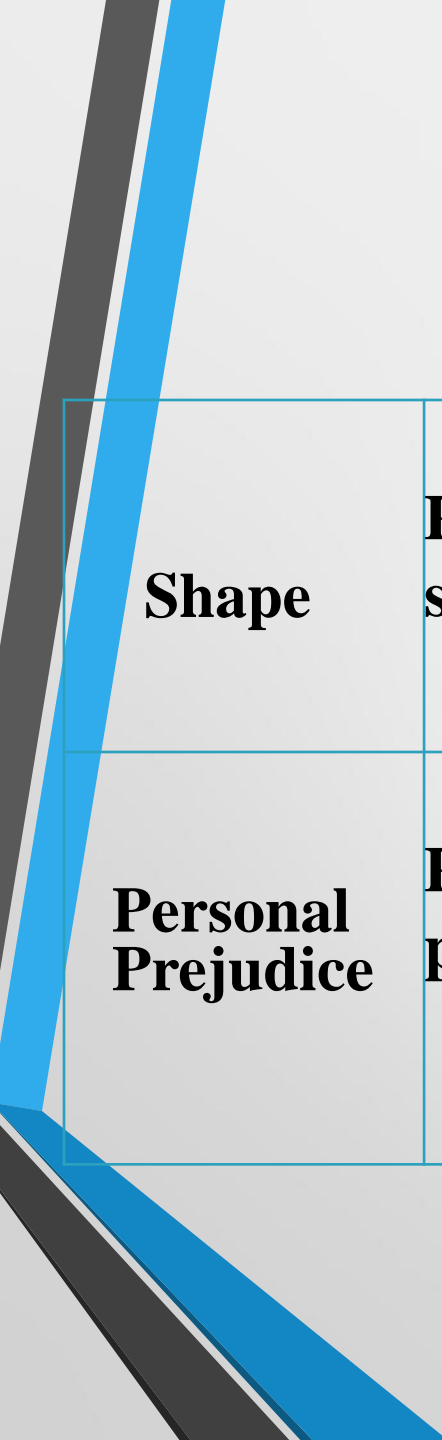
Primary data are expensive and time consuming.

Secondary data are relatively cheaper and less time consuming for data collection.

Efforts

More efforts required for collecting primary data.

Comparatively secondary data collection required less effort.



| | | |
|---------------------------|--|--|
| Shape | Primary data are in the shape of raw material. | Secondary data are usually in the shape of ready made products. |
| Personal Prejudice | Possibilities of personal prejudice are higher. | There are lesser possibilities of personal prejudice. |

SOURCES/METHODS OF DATA COLLECTION

```
graph TD; A[SOURCES/METHODS OF DATA COLLECTION] --> B[PRIMARY DATA]; A --> C[SECONDARY DATA]; B --> D[OBSERVATION]; B --> E[INTERVIEW]; B --> F[QUESTIONNAIRE]; B --> G[SCHEDULE]; E --> H[SURVEY]; F --> I[PANEL METHOD]; F --> J[CASE STUDY METHOD]; C --> K[PUBLIC DOCUMENTS]; C --> L[PRIVATE / PERSONAL DOCUMENTS];
```

The diagram is a hierarchical flowchart titled 'SOURCES/METHODS OF DATA COLLECTION'. It branches into two main categories: 'PRIMARY DATA' and 'SECONDARY DATA'. 'PRIMARY DATA' further branches into 'OBSERVATION', 'INTERVIEW', 'QUESTIONNAIRE', and 'SCHEDULE'. 'INTERVIEW' and 'QUESTIONNAIRE' both lead to 'SURVEY'. 'QUESTIONNAIRE' also leads to 'PANEL METHOD'. 'PANEL METHOD' leads to 'CASE STUDY METHOD'. 'SECONDARY DATA' branches into 'PUBLIC DOCUMENTS' and 'PRIVATE / PERSONAL DOCUMENTS'. The boxes are light blue with black text and are connected by black lines.

PRIMARY DATA

OBSERVATION

INTERVIEW

QUESTIONNAIRE

SCHEDULE

SURVEY

PANEL METHOD

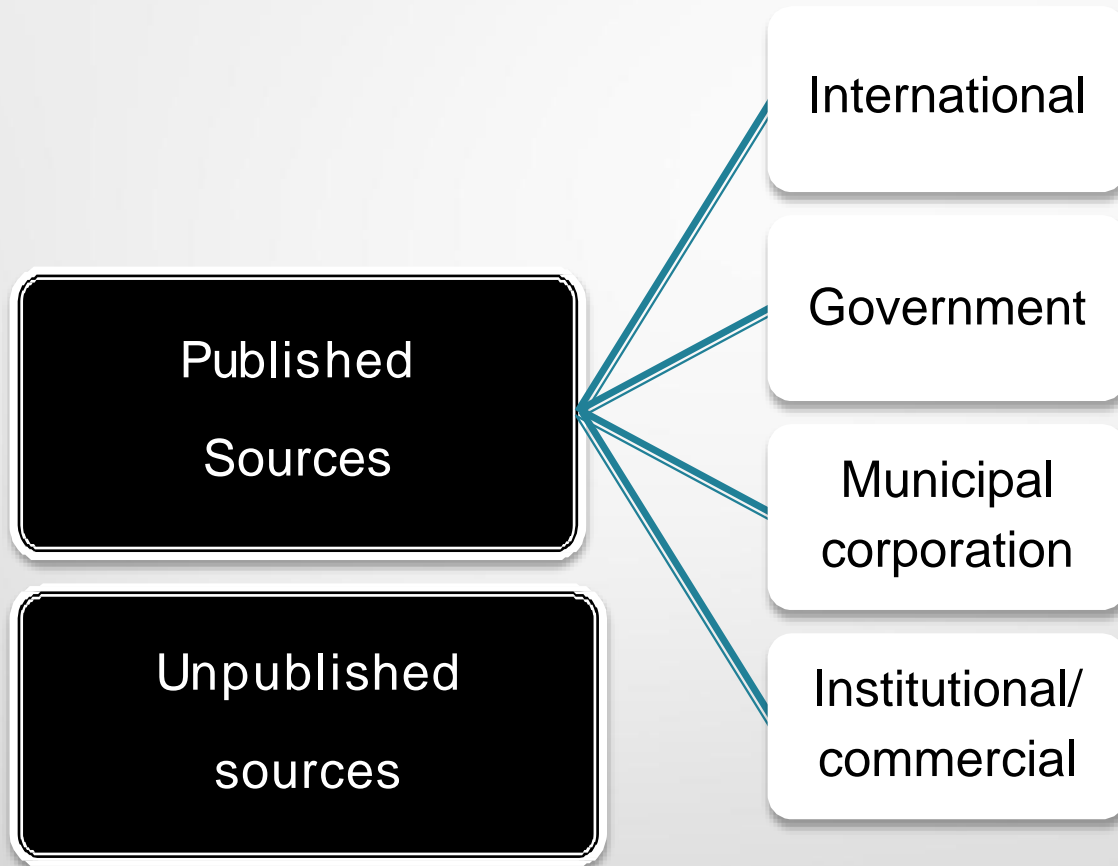
CASE STUDY METHOD

SECONDARY DATA

PUBLIC DOCUMENTS

PRIVATE / PERSONAL DOCUMENTS

Method of Collection Secondary Data



SOURCES/METHODS OF PRIMARY DATA **COLLECTION: -**

OBSERVATION METHOD:

A method under which data from the field is collected with the help of observation by the observer.

No conversation or communication should be done while observing.

“Observation may be defined as systematic viewing, coupled with consideration of seen phenomenon.”

P.V. Young

TYPES OF OBSERVATION:-

➤ **Structured**

Merits:

- **Filtering what is relevant and focusing on it.**
- **Avoiding waste of time.**
- **Studying only the target group.**

Demerits:

- **limiting the independence of the observer, crucial things might be left out from being observed.**

➤ **Unstructured Observation**

Merits:

- **everything is taken note of and the researcher then segregates the related and relevant details.**

Demerits:

- **experienced**

TYPES OF OBSERVATION:-

➤ **Participant**

Merits:

- **The observer will not miss anything which is relevant.**
- **There is very scope for him to clarify things which he cannot understand.**
- **He can interact with people to get more information.**

Demerits:

- **The conduct or behaviour or reaction of the target audience may be influenced by his presence.**
- **He tends to focus more on things of interest to him, than what is relevant.**

Non-Participant Observation

Strength of this method:

Observer would collect firsthand information without being noticed or influenced by anyone.

ADVANTAGES:

- **It does not rely on people's willingness to provide information.**
- **Collect data where and when an event or activity is occurring.**
- **This method can be used with interview, survey, and some other data collection method.**
- **No need of question preparation.**
- **It is suitable to studies those respondents who are not capable of giving verbal response.**

DISADVANTAGES:

- people usually perform better when they know they are being observed.
- It is a time-consuming method.
- Limited information.
- Lack of verbal communication causes barriers.
- Depends on observer own qualities.
- This method is suitable for smaller setting.

2) INTERVIEW METHOD: This method of collecting data involves presentation or oral-verbal and reply in terms of oral-verbal responses. Conversation and communication is the main tool of interview.

Types of Interviews:

A) Personal interviews
C) Structured interviews

B) Telephonic interviews
D) Unstructured interviews

TYPES OF INTERVIEWS:

A) Structured interviews: In this case, a set of pre- decided questions are there.

Merits:

- Different respondents have different type of answers to the same structure of questions.
- They can be used to get in touch with a large sample of the target population.
- asking accurate research questions.
- often generates reliable results and is quick to execute.
- relationship between the researcher and the respondent

Demerits:

- Limited scope of assessment of obtained results.
- The accuracy of information overpowers the detail of information.
- Respondents are forced to select from the provided answer options.

TYPES OF INTERVIEWS:

Unstructured interviews: In this case, we don't follow a system of pre-determined questions.

Merits:

- extremely easy for researchers
- The participants and researcher can clarify all their doubts about the questions for better answers.

Demerits:

- researchers take time to execute these interviews.
- reliability of unstructured interviews is questionable
- the ethics involved in these interviews are considered borderline upsetting.

TYPES OF INTERVIEWS:

Personal interviews: The interviewer asks questions generally in a face to face contact to the other person or persons.

Merits:

- **Higher response rate.**
- **Easy way to adapt the questions if this is not understood.**
- **More complete answers can be obtained**
- **body language at the time of asking the questions and taking notes about it.**

Demerits:

- **They are time-consuming and extremely expensive.**
- **They can generate distrust on the part of the interviewee.**

TYPES OF INTERVIEWS:

Telephonic interviews: When it is not possible to contact the respondent directly, then interview is conducted through – Telephone.

Advantages:

- They are usually lower cost.
- The information is collected quickly.
- Having a personal contact can also clarify doubts or give more details of the questions.

Disadvantages:

- phone calls are not answered.
- Good communication is vital for the generation of better answers.

ADVANTAGES:

- **Interviewer can collect supplementary information about respondent's personal characteristics and environment which has value in interpreting results.**
- **Non Responses are generally low.**
- **Samples can be controlled more effectively.**
- **Allows respondents to describe what is important to them.**
- **There is greater flexibility; due to restructuring questions can be done.**

DISADVANTAGES:

- **Respondent may give bias information.**
- **Expensive method.**
- **Some Executive people are not approachable so data collected may be inadequate.**
- **Takes more time when samples are more, Systematic errors may be occurred.**

3) QUESTIONNAIRE METHOD: This is a set of questions arranged logically, divided into groups, with the object of collecting information for research. The questionnaire is mailed to respondents who are expected to read and understand the questions and write down the reply in the space meant for the purpose in the questionnaire itself or chooses the reply among all choices available on closed end questionnaire.

VARIOUS FORMS OF QUESTIONS USED IN QUESTIONNAIRE:

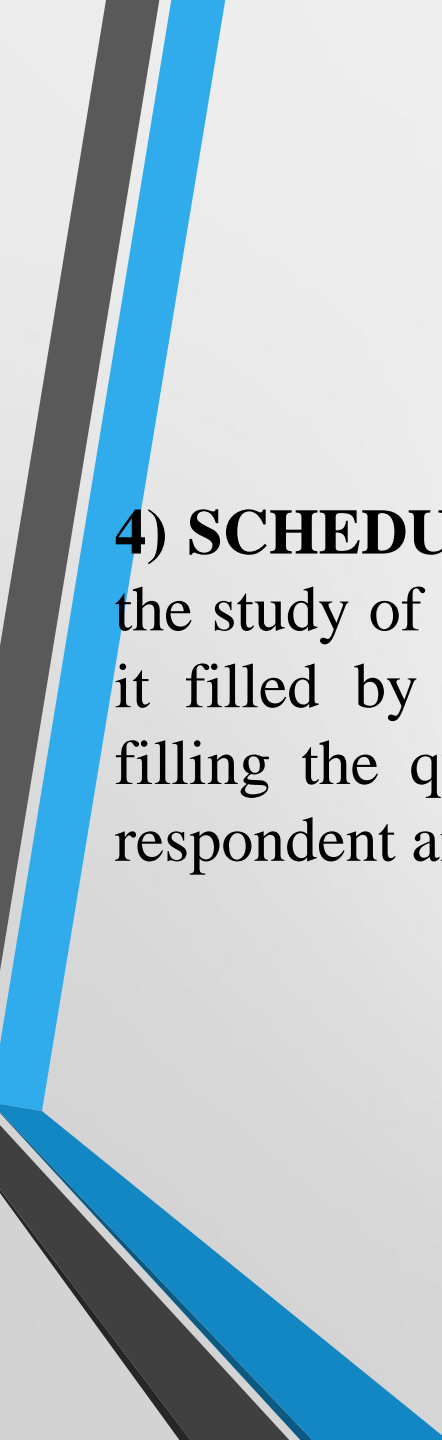
- A) Direct Question and Indirect Question
- B) Open Form of Questions and Closed-End Form of Questions
- C) Multiple-Choice Questions (MCQ) and Scale or Rating Questions .

ADVANTAGES:-

- **Low cost even the geographical area is large to cover.**
- **Answers are in respondent's own words so free from bias.**
- **Adequate time to think for answers.**
- **Non approachable respondents may be conveniently contacted.**
- **Large samples can be used so results are more reliable.**

DISADVANTAGES:

- **Time consuming and expensive compared to other data collection methods.**
- **It is very expensive method.**
- **Low rate of return of duly filled questionnaire.**
- **It can be used when respondent is educated and co-operative.**
- **Difficult to know the expected respondent have filled the form or it is filled by some one else.**



4) SCHEDULE METHOD: It is one of the important methods for the study of social problems. Schedules is like a questionnaires, but it filled by enumerator. Enumerators are specially appointed for filling the questionnaire, they explains the aim and objective to respondent and fill the answers in provided space.

ADVANTAGES:

- **In Schedule the information is collected complete and accurate.**
- **Direct personal contact is established with respondents and useful in extensive enquiries.**
- **Population census all over the world is conducted through this method.**
- **Non response is low.**

DISADVANTAGES:

- **Depends on Honesty and competence of enumerator.**
- **It's a very expensive and time-consuming method.**
- **Not very useful for small organization or small budget research purpose.**

5) SURVEY METHOD: One of the common methods of diagnosing and solving of social problems are that of undertaking surveys.

Surveys are....

- 1.A detailed inspection or investigation.
- 2.A general or comprehensive view.
- 3.A gathering of a sample of data or opinions considered to be representative of a whole.

TYPES OF SURVEY:-

- **General or Specific survey**
- **Census or sample survey**
- **Public opinion surveys**
- **Private surveys**
- **Confidential survey**
- **Postal survey**

ADVANTAGES:

- **Valid and reliable conclusion.**
- **Helpful in the progress of science.**
- **Study of social changes and problems.**
- **Some people feel more comfortable responding to a survey than participating in an interview.**

DISADVANTAGES:

- **Good surveys are time consuming.**
- **Given lack of contact with respondent, never know who really completed the survey.**
- **Size and diversity of sample will be limited by people's ability to read.**
- **Survey respondents may not complete the survey resulting in low response rates.**

6) PANEL METHOD: In this method, data is collected **interviewing**
the same group of people on two or more occasions.

This is used for studies on:

- 1)Expenditure Pattern
- 2)Consumer Behavior
- 3)Effectiveness of Advertising
- 4)Voting Behavior and so on

ADVANTAGES: -

- Best method for marketing research.
- Helpful in discovering latest changes on trends.
- This method is useful before the introduction of any product and after that too.
- Helpful in determining taste and preference of consumer.

DISADVANTAGES:-

- Time consuming.
- Useful in long and detailed research only.
- Expensive method.
- Depends on respondents honesty and knowledge.

7) CASE STUDY METHOD: It is an appropriate tool of data collection in studying an individual a family an institution or group behavior in detail. It is essentially an intensive investigation of the unit under consideration.

Its important characteristics are as follows:

- a)The researcher can take one single social unit or more of such units for his study purpose.
- b)The selected unit is studied intensively i.e. it is studied in minute details.
- c)The behavior pattern of the concerning unit is studied.

ADVANTAGES:

- It is helpful in theory building & testing.
- It is widely used in the studies of psychology, industry, & for anthropological research.
- Fully depicts people's experience in program input, process, and results.
- It helps in formulating relevant hypotheses along with the data which may be helpful in testing them.
- Helps to construct appropriate questionnaire or schedule.

DISADVANTAGES:

- From a methodological view point it is improbable to classify data into a uniform order.
- Case data is hardly comparable with statistical quantitative data.
- Classification & generalization is never being done.
- It's based on several assumptions which may not be very realistic.
- It can use in a limited sphere, not for big society case, sampling is also not possible.

METHODS/SOURCES OF SECONDARY DATA COLLECTION:-

Secondary data are available mainly in two forms- published data and unpublished data:-

PUBLISHED DATA: Published data are more often the information required by individuals and organizations is published in some form or the other in consideration of user's need. It can used in annexure too.

e.g. the companies publish their financial statements in the form of quarterly or half yearly or annual reports.

UNPUBLISHED DATA: - Unpublished data are that secondary information which is available from records which are not published due to privacy or variety of reasons.

e.g. works of scholars, research workers, trade associations etc.

Secondary sources can be divided into two kinds:-

1) PERSONAL/PRIVATE RECORDS OR DOCUMENTS-

This documents denotes individual's feelings, opinions and an idea about different socio-cultural changes, social incidents and structural changes. These records can be in both published and unpublished form.

A) Letters,

B) Personal diaries,

C) Autobiography and memories.

D) Enquiries or investigation of private nature for use of their members only.

Unpublished Personal Records

- Diaries
- Letters

Government Records

- Census Data/population statistics:
- Health records
- Educational institutes records

Public Sector Records:

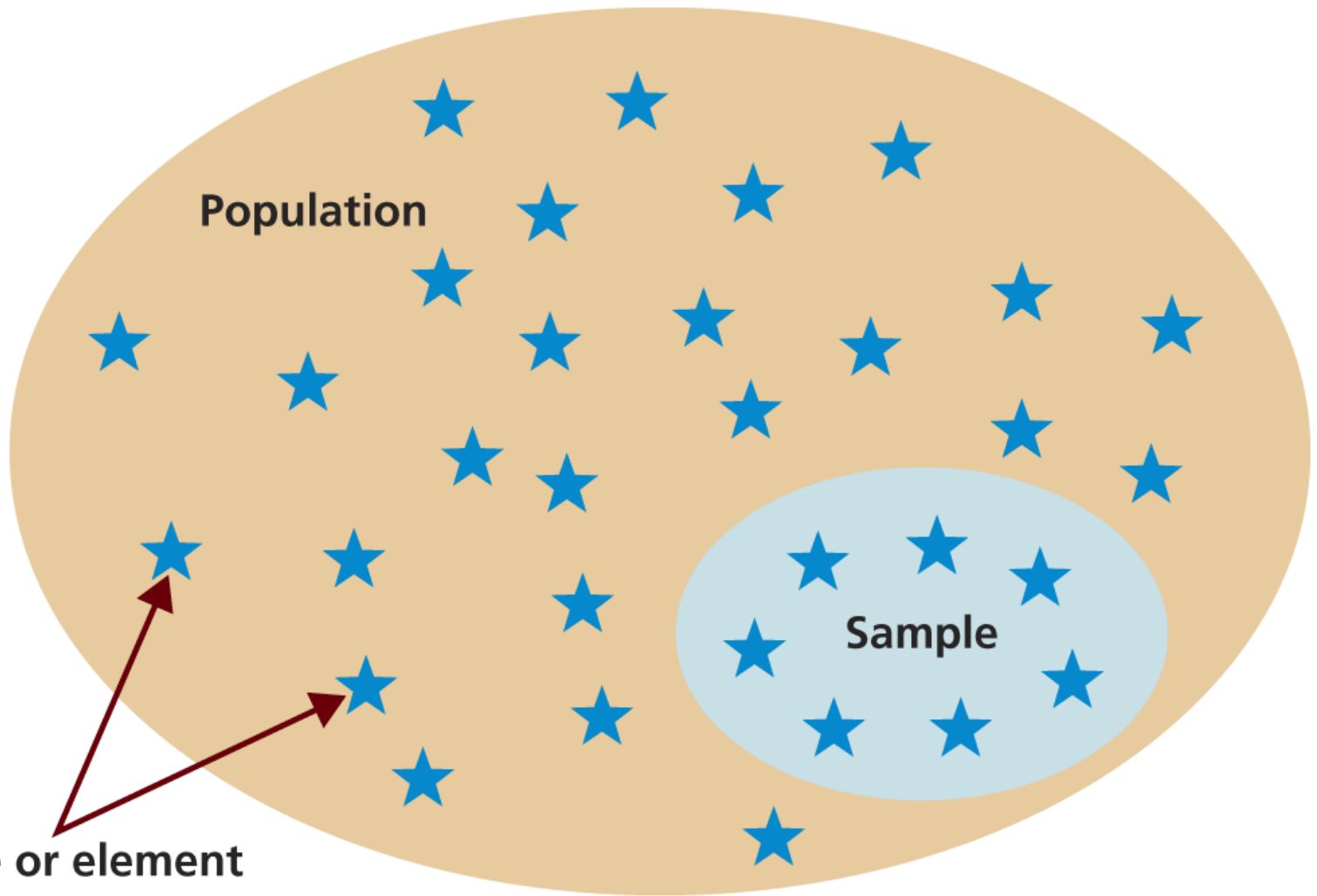
- NGOs's survey data
- Other private companies records



SAMPLING and SAMPLING METHODS

SAMPLING

- **This process of selecting just a small group of cases from out of a large group is called sampling.**
- **A process or method of drawing a representative group of individuals or cases from a particular population.**



Population

Sample

Case or element

The need to sample

Sampling- a valid alternative to a census when:

- A survey of the entire population is impracticable.
- Budget constraints restrict data collection.
- Time constraints restrict data collection.
- Results from data collection are needed quickly.

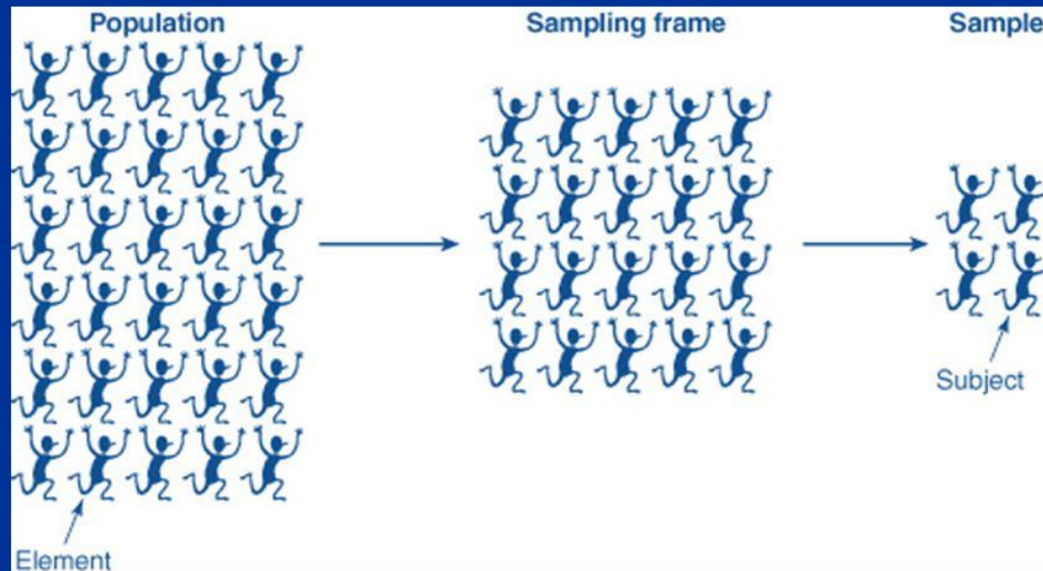
Population in Research

- It does not necessarily mean a number of people, it is a collective term used to describe the total quantity of things (or cases) of the type which are the subject of your study.
- So a *population* can consist of certain types of objects, organizations, people or even events.

Sampling Frame

- Within this population, there will probably be only certain groups that will be of interest to your study, this *selected category* is your sampling frame.

Relationship between Population, Sampling Frame and Sample

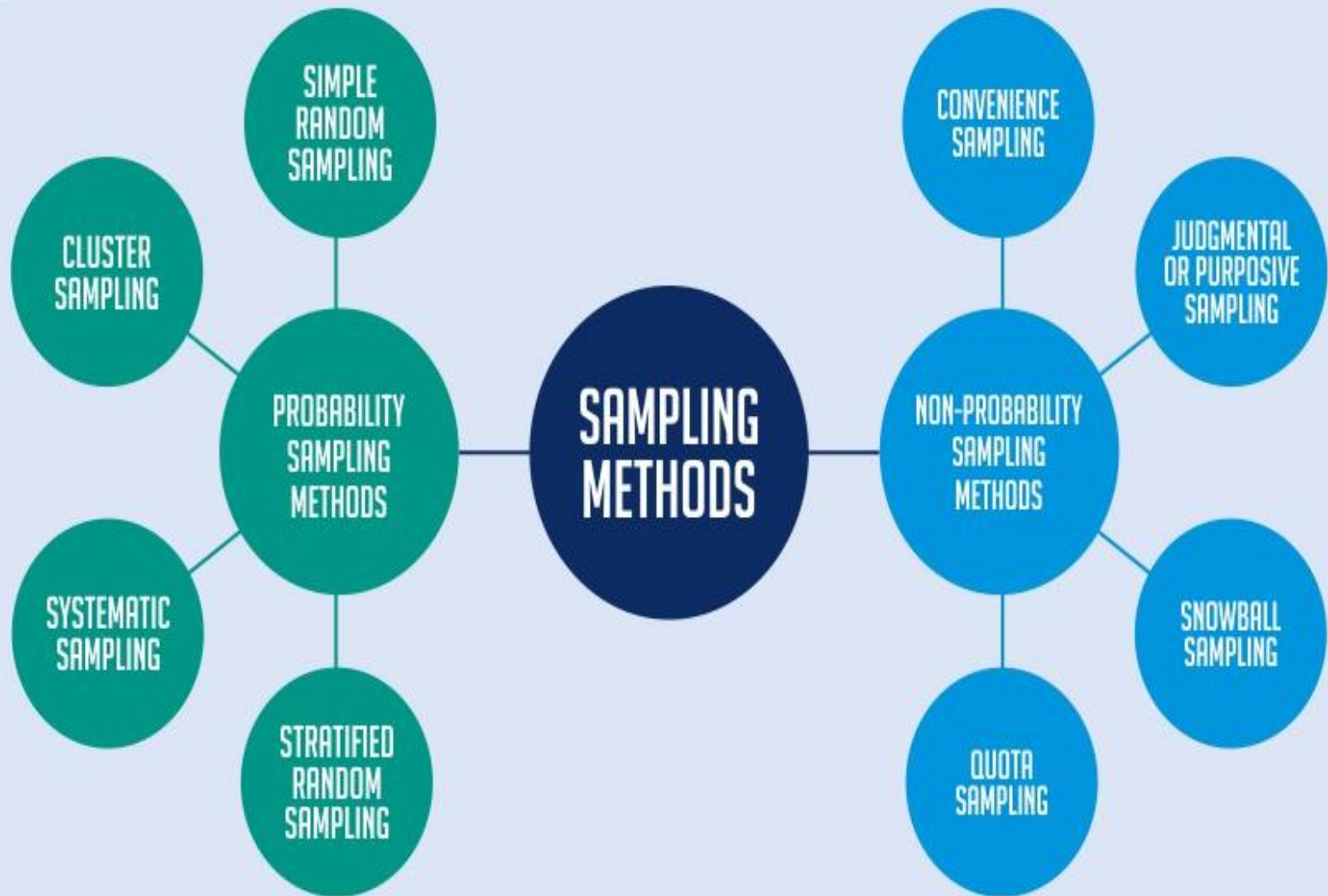


Populations can have the following characteristics:

| Characteristics | Explains | Examples |
|--------------------------------|--|---|
| homogeneous | all cases are similar | bottles of water on a production line |
| stratified | contain strata or layers | people with different levels of income: low, medium, high |
| proportional stratified | contains strata of known proportions | percentages of different nationalities of students in a university |
| grouped by type | contains distinctive groups | of apartment buildings – towers, slabs, villas, tenement blocks |
| grouped by location | different groups according to where they are | animals in different habitats – desert, equatorial forest, savannah, tundra |



SAMPLING METHODS



- ✓ **Probability sampling** techniques give the most reliable representation of the whole population.
- ✓ **Non-probability techniques**, relying on the judgment of the researcher or on accident, cannot generally be used to make generalizations about the whole population.



Probability Sampling

- It is a sampling technique in which sample from a larger population are chosen using a method based on the theory of probability.
- For a participant to be considered as a probability sample, he/she must be selected using a random selection.
- The most important requirement of probability sampling is that everyone in your population has a known and an equal chance of getting selected.
- Probability sampling uses statistical theory to select randomly, a small group of people (sample) from an existing large population and then predict that all their responses together will match the overall population.

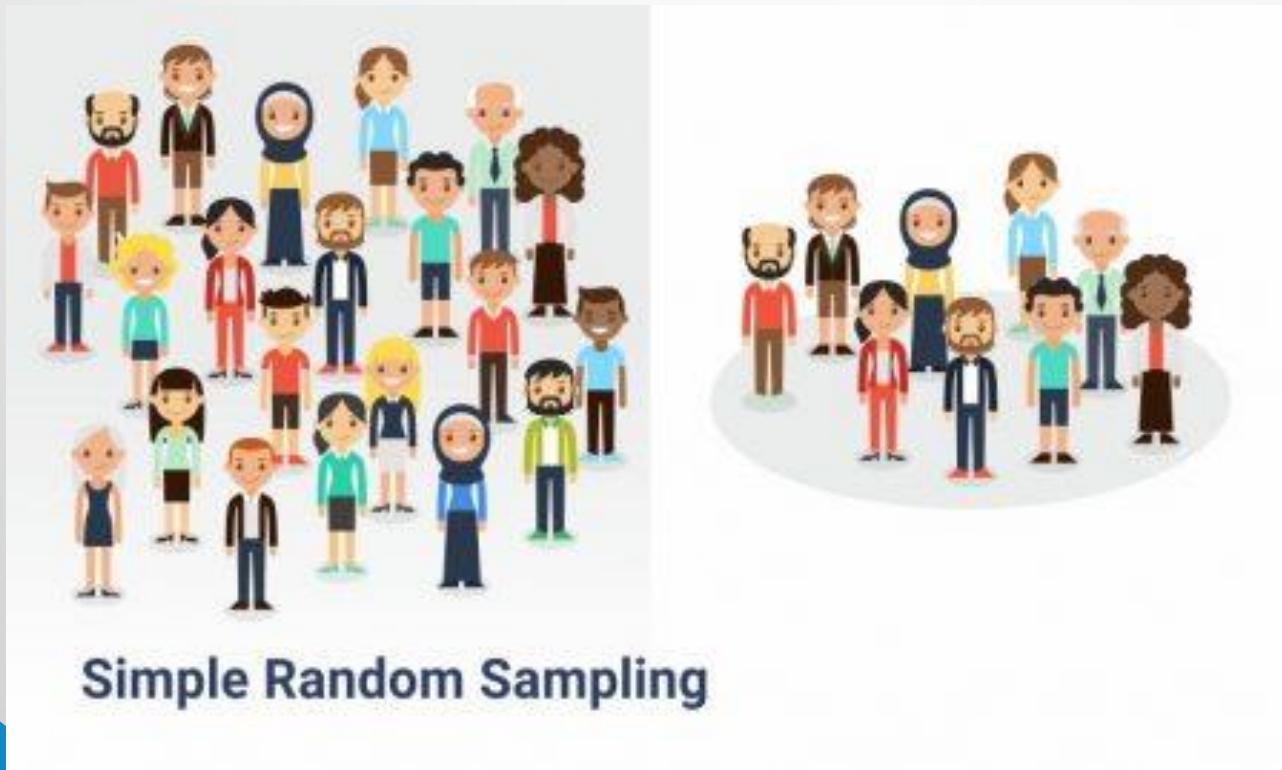
Types of Probability Sampling

Four main techniques used for a probability sample:

- Simple random Sampling
- Stratified random Sampling
- Cluster Sampling
- Systematic Sampling

Simple Random Sampling (S.R.S)

- **Simple random sampling (S.R.S.) is the technique in which sample is so drawn that each unit in the population has an equal and independent chance of being included in the sample.**



PROCEDURE OF SELECTING A RANDOM SAMPLE

The first step is to prepare a list of all (N) units in the population and number them serially from **1 to N** .

Then a sample of n units is taken by are the following methods.

(a) Lottery method

(b) Use of random numbers tables

i) Remainder approach

ii) Quotient approach

The samples can be drawn in two possible ways.

**1) SIMPLE RANDOM SAMPLING WITHOUT REPLACEMENT
(srswor) :**

If the unit selected in any draw is not replaced in the population before making the next draw, then it is known as **simple random sampling without replacement (srswor)**

The n units of the sample are drawn from the population one by one, the unit obtained at any drawn not being replaced in the population, in such a way the probability of any unit in the first draw is $1/N$, that of any unit in the second drawn is $1/(N-1)$,..... that of any unit in the n^{th} draw is $1/(N-n+1)$ and so on.

Note that a unit can be selected at any one of the n draws.

Let u_i be the i^{th} unit selected in the sample. This unit can be selected in the sample either at first draw, second draw, ..., or n^{th} draw.

Let $P_j(i)$ denotes the probability of selection of u_i at the j^{th} draw, $j = 1, 2, \dots, n$.

Then

$$\begin{aligned} P_j(i) &= P_1(i) + P_2(i) + \dots + P_n(i) \\ &= 1/N + 1/N + 1/N \text{ (n times)} \\ &= n/N \end{aligned}$$

SIMPLE RANDOM SAMPLING WITH REPLACEMENT (SRSWR) :

The sampling units are chosen with replacement because the selected units are placed back in the population.

the n units of the sample are drawn from the population one by one, the units obtained at any draw being replaced in the population, in such a way that the probability of drawing any unit in any draw is $1/N$.

When n units are selected with SRSWR, the total number of possible samples are . N^n

The Probability of drawing a sample is $1/N^n$

Notations:

The following notations will be used in further notes:

N : Number of sampling units in the population (Population size).

n : Number of sampling units in the sample (sample size)

Y : The characteristic under consideration

Y_i : Value of the characteristic for the i^{th} unit of the population

$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$: sample mean

$\bar{Y} = \frac{1}{N} \sum_{i=1}^N Y_i$: population mean

$$S^2 = \frac{1}{N-1} \sum_{i=1}^N (Y_i - \bar{Y})^2 = \frac{1}{N-1} \left(\sum_{i=1}^N Y_i^2 - N\bar{Y}^2 \right)$$

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N (Y_i - \bar{Y})^2 = \frac{1}{N} \left(\sum_{i=1}^N Y_i^2 - N\bar{Y}^2 \right)$$

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2 = \frac{1}{n-1} \left(\sum_{i=1}^n y_i^2 - n\bar{y}^2 \right)$$

Variance of Sample Mean

Variance of sample mean under SRSWOR

$$V(\bar{y}_{WOR}) = E(\bar{y} - \bar{Y})^2 = \frac{N-n}{Nn} S^2$$

Variance of sample mean under SRSWR

$$V(\bar{y}_{WR}) = E(\bar{y} - \bar{Y})^2 = \frac{N-1}{Nn} S^2$$

Example 1:

In a population with $N=5$, the values of y are 7,1,10,3,9. calculate the sample mean \bar{y} and sample variance s^2 for all the simple random samples(SRSWOR) of size 2 and verify the following.

i) $E(\bar{y}) = \bar{Y}$

ii) $E(s^2) = S^2$

iii) $\text{var}(\bar{y}) = \frac{N-n}{nN} S^2$

iv) $\text{Var}(\text{SRSWR}) \geq \text{Var}(\text{SRSWOR})$

Example 2:

Suppose that a population consists of 6 units with measurements 2, 4, 6, 8, 10 and 12.

Write all the possible samples of size 2 by without replacement from the population and verify

- i) The sample mean is unbiased estimate of the population
 - ii) The sample mean square is unbiased estimate if the population mean square
 - iii) Also calculate the sampling variance of the estimate, sample mean and verify it with the formula of variance
- SRSWOR is efficient than SRSWR and find the gain in efficiency.

Stratified Random Sampling

The procedure of partitioning a given population into homogeneous groups is called strata and then selecting samples independently from each stratum is known as stratified sampling, if a sample from each stratum is selected by random sampling, the procedure will be called stratified random sampling.

Strata are constructed such that they are

Within homogeneous

Among heterogeneous

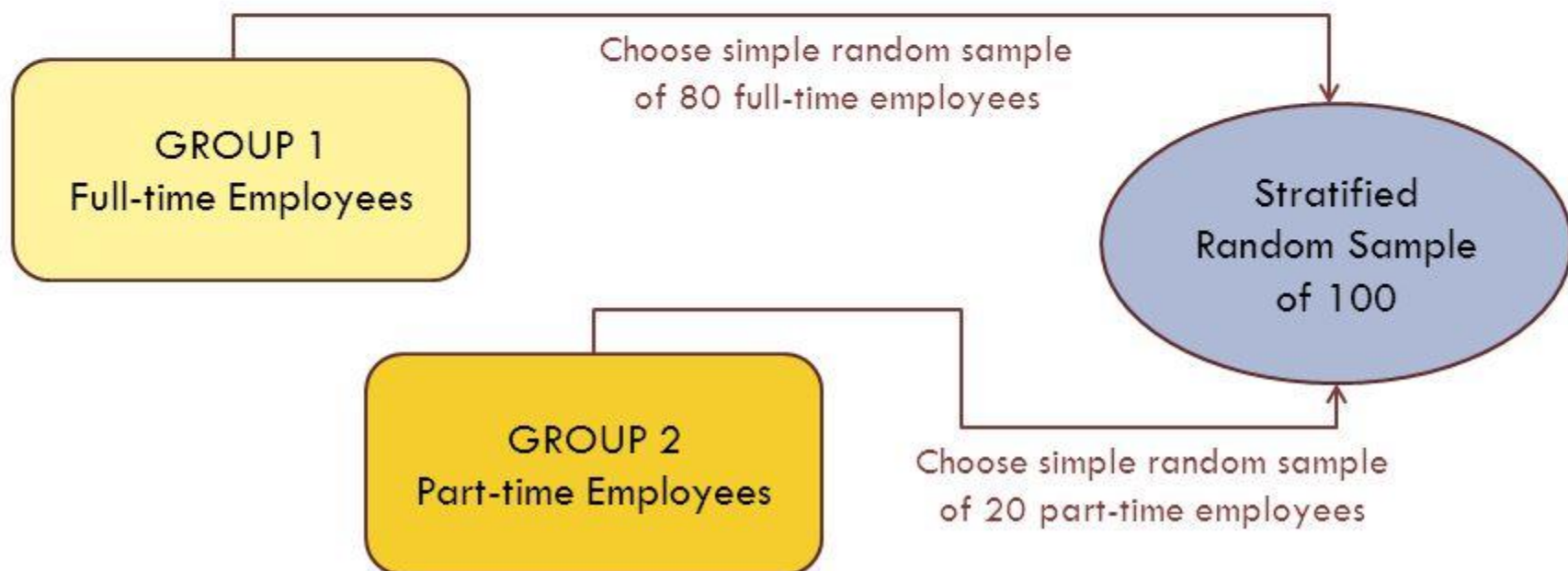


Stratified Random Sampling



Example – Stratified Random Sampling

A company has 800 full-time and 200 part-time employees. To draw a sample of 100 employees, a simple random sample of 80 full-time employees is selected and a simple random sample of 20 part-time employees is selected.



Merits:

- **methods covers diverse characteristics of the population.**
- **Comparative analysis of the data becomes possible.**
- **Reliable and meaningful results.**

Demerits:

- **Need to have complete knowledge.**
- **Bias at the time of classification.**
- **If population size is small, difficult to further divide it into smaller strata**

Notations

| | |
|--|---|
| N_h (size of the h th stratum) | total number of units in the h th stratum; |
| n_h | number of units selected in the sample from the h th stratum; |
| $W_h = \frac{N_h}{N}$ (h th stratum weight) | proportion of the population units falling in the h th stratum; |
| $f_h = \frac{n_h}{N}$ | sampling fraction for the h th stratum; |
| Y_{hi} (y_{hi}) | the value of study variable for the i th unit (or sample unit) in the h th stratum, $1 \leq i \leq N_h$; |
| $Y_h = \sum_{i=1}^{N_h} Y_{hi}$ | h th stratum total for the estimation variable based on N_h units; |
| $\bar{Y}_h = \frac{1}{N_h} \sum_{i=1}^{N_h} Y_{hi}$ | mean of the study variable in the h th stratum; |
| $\bar{y}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} Y_{hi}$ | h th stratum sample mean for the study variable; |
| $\sigma_h^2 = \frac{1}{N_h} \left(\left(\sum_{i=1}^{N_h} Y_{hi}^2 \right) - N_h \bar{Y}_h^2 \right)$ | h th stratum variance based on N_h units; |
| $S_h^2 = \frac{N_h}{N_h - 1} \sigma_h^2$ | h th stratum mean square based on N_h units; |
| $s_h^2 = \frac{1}{n_h - 1} \left(\left(\sum_{i=1}^{n_h} y_{hi}^2 \right) - n_h \bar{y}_h^2 \right)$ | sample mean square based on n_h sample units drawn from the h th stratum; |

EXAMPLE

The data given below pertains to the total geographical area in 20 villages of a block. Treating this as population of 20 units, stratify this population in three strata taking stratum sizes to be villages with geographical area, 50 ha. or less, villages with geographical area more than 50 ha. & up to 100 ha. and villages having geographical area more than 100 ha. A sample of 6 villages is to be selected by taking 2 villages in each stratum. Compare the efficiency of stratified sampling with corresponding simple random sampling.

Population(N)=

[20,80,50,100,150,70,20,250,220,10,50,140,80,20,50,30,70,90,100,120]

Cluster Random Sampling

- It is a way to randomly select participants when they are geographically spread out. Cluster sampling usually analyzes a particular population in which the sample consists of more than a few elements, for example, city, family, university etc. The clusters are then selected by dividing the greater population into various smaller sections.



Stratified & Cluster Sampling

Stratified

- Population divided into **few** subgroups
 - Each subgroup has **many** elements in it.
 - Subgroups are selected according to some criterion that is related to the variables under study.
- **Homogeneity** within subgroups
- **Heterogeneity** between subgroups
- Choice of **elements** from within each subgroup

Cluster

- Population divided into **many** subgroups
 - Each subgroup **few** elements in it.
 - Subgroups are selected according to some criterion of ease or availability in data collection.
- **Heterogeneity** within subgroups
- **Homogeneity** between subgroups
- Random choice of **subgroups**

Systematic Sampling

- It is when you choose every “ n^{th} ” individual to be a part of the sample. For example, you can choose every 5th person to be in the sample. Systematic sampling is an extended implementation of the same old probability technique in which each member of the group is selected at regular periods to form a sample. There’s an equal opportunity for every member of a population to be selected using this sampling technique.



Types of Non-probability Sampling

Four main techniques used for a non-probability sample:



Convenience



Judgemental



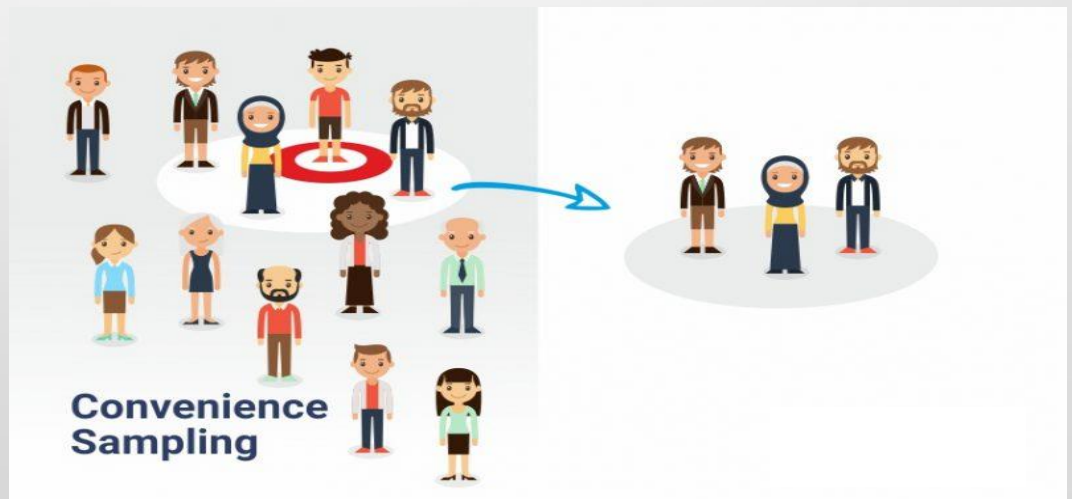
Snowball



Quota

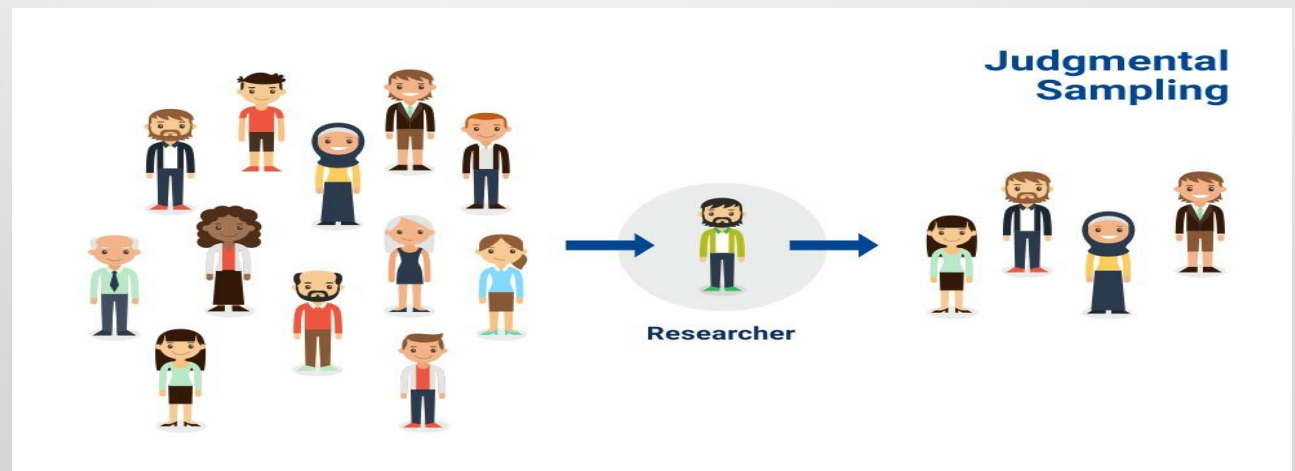
Convenience Sampling

- It is a non-probability sampling technique used to create sample as per ease of access, readiness to be a part of the sample, availability at a given time slot or any other practical specifications of a particular element.
- Convenience sampling involves selecting haphazardly those cases that are easiest to obtain for your sample, such as the person interviewed at random in a shopping center for a television program.



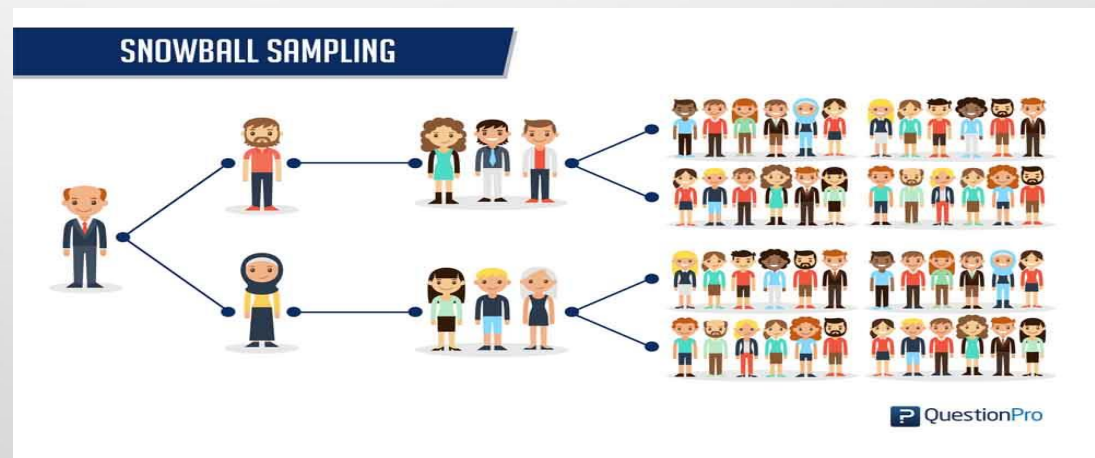
Judgmental Sampling

- In the judgmental sampling, also called purposive sampling, the sample members are chosen only on the basis of the researcher's knowledge and judgment.
- It enables you to select cases that will best enable you to answer your research question(s) and to meet your objectives.



Snowball Sampling

- Snowball sampling method is purely based on referrals and that is how a researcher is able to generate a sample. Therefore this method is also called the chain-referral sampling method.
- This sampling technique can go on and on, just like a snowball increasing in size (in this case the sample size) till the time a researcher has enough data to analyze, to draw conclusive results that can help an organization make informed decisions.



Quota Sampling

- Selection of members in this sampling technique happens on basis of a pre-set standard. In this case, as a sample is formed on basis of specific attributes, the created sample will have the same attributes that are found in the total population. It is an extremely quick method of collecting samples.
- Quota sampling is therefore a type of stratified sample in which selection of cases within strata is entirely non-random.

