RESEARCH METHODOLOGY

Research Design

Dr.Roopa Ravish
Dept of Computer Science and Engineering

RESEARCH METHODOLOGY

Research Design

Department of Computer Science and Engineering

Contents – Chapter 3

MEANING OF RESEARCH DESIGN

Research Design is the preparation of the design of the research project. It constitutes the blueprint for the collection, measurement and analysis of data.

Design decisions revolve around the following questions:

1. What is the study about?
2. Why is the study being made?
3. Where will the study be carried out?
4. What type of data is required?
5. Where can the required data be found?
6. What periods of time will the study include?
7. What will be the sample design?
8. What techniques of data collection will be used?
9. How will the data be analyzed?
10. In what style will the report be prepared?

RESEARCH DESIGN-BREAK DOWN

1. The sampling design
   - Deals with the method of selecting items to be observed for the given study.
2. The observational design
   - Relates to the conditions under which the observations are to be made.
3. The statistical design
   - Concerns with the question of how many items are to be observed and how the information and data gathered are to be analyzed.
4. The operational design
   - Deals with the techniques by which the procedures specified in the sampling, statistical and observational designs can be carried out.

NEED FOR RESEARCH DESIGN

1. It facilitates the smooth sailing of the various research operations.
2. Making research as efficient as possible yielding maximal information with minimal expenditure of effort, time and money.
3. We need a research design or a plan in advance of data collection and analysis for our research project.
4. Research design stands for
   - advance planning of the methods to be adopted for collecting the relevant data and
   - the techniques to be used in their analysis,
   - keeping in view the objective of the research and the availability of staff, time and money.

FEATURES OF A GOOD DESIGN

- A good design is often characterised by adjectives like flexible, appropriate, efficient, economical and so on.
- Generally, the design which minimises bias and maximises the reliability of the data collected and analysed is considered a good design.
- The design which gives the smallest experimental error is supposed to be the best design in many investigations.
- A design which yields maximal information and provides an opportunity for considering many different aspects of a problem is considered most appropriate and efficient design in respect of many research problems.

Thus, the question of good design is related to the purpose or objective of the research problem and also with the nature of the problem to be studied. A design may be quite suitable in one case, but may be found wanting in one respect or the other in the context of some other research problem. One single design cannot serve the purpose of all types of research problems.

FEATURES OF A GOOD DESIGN

A research design appropriate for a particular research problem, usually involves the consideration of the following factors:

(i) the means of obtaining information;
(ii) the availability and skills of the researcher and his staff, if any;
(iii) the objective of the problem to be studied;
(iv) the nature of the problem to be studied; and
(v) the availability of time and money for the research work

IMPORTANT CONCEPTS RELATING TO RESEARCH DESIGN

1. Dependent and Independent Variables
2. Extraneous Variable
3. Control
4. Confounded Relationship
5. Research Hypothesis
6. Experimental and Non-experimental Hypothesis- Testing Research
7. Experimental and Control Groups
8. Treatments
9. Experiment
10. Experimental Unit(s)

IMPORTANT CONCEPTS RELATING TO RESEARCH DESIGN

Dependent and Independent Variable

Variable: A concept which can take on different quantitative values.

For example, concepts like weight, height, income, etc.

Continuous variables – phenomena which can take on quantitatively different values even in decimal points.

For example, age.

Discontinuous or Discrete variables – If some variables can only be expressed in integer values.

For example, no. of children.

If one variable depends upon or is a consequence of the other variable, it is termed as dependent variable, and the variable that is *antecedent* to the dependent variable is termed as independent variable.

IMPORTANT CONCEPTS RELATING TO RESEARCH DESIGN

Dependent and Independent Variable

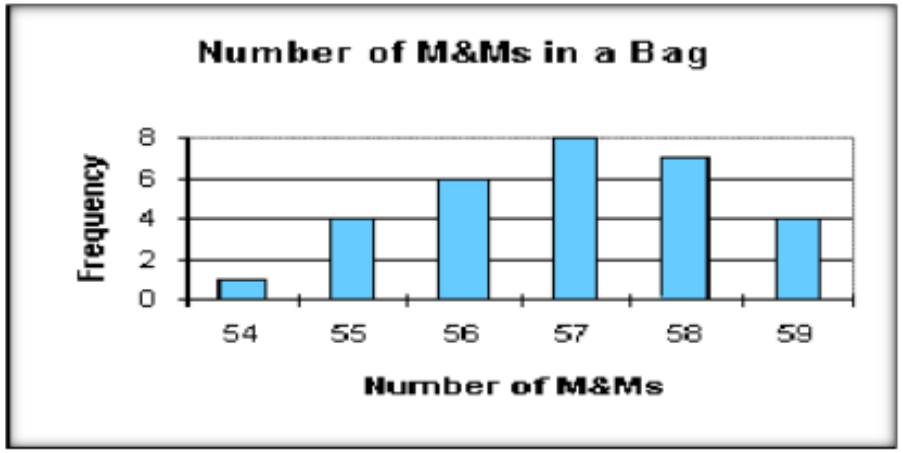
CONTINUOUS VARIABLE

Phenomena which can take on quantitatively different values even in decimal points .

AGE IS AN EXAMPLE

IMPORTANT CONCEPTS RELATING TO RESEARCH DESIGN

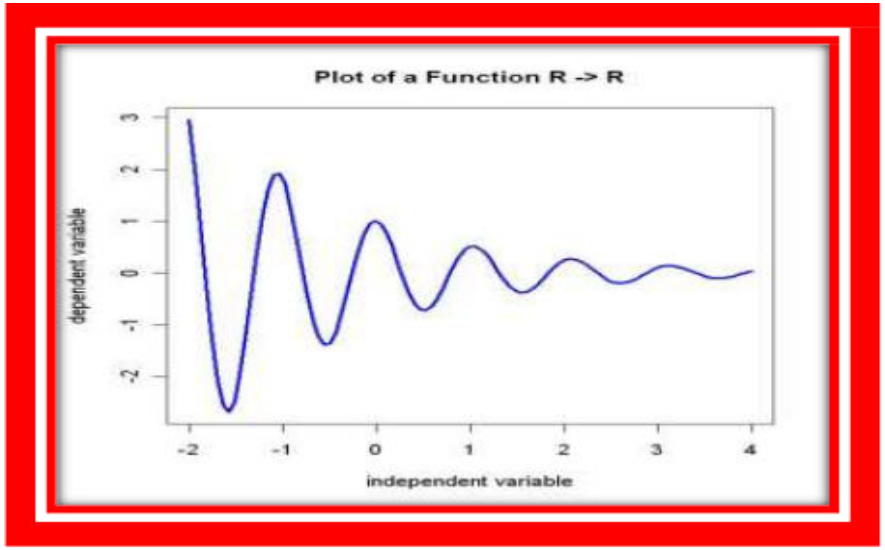
DISCRETE VARIABLE



All variables are not continuous.
Only if they are expressed in integer values they are non continuous variables.

IMPORTANT CONCEPTS RELATING TO RESEARCH DESIGN

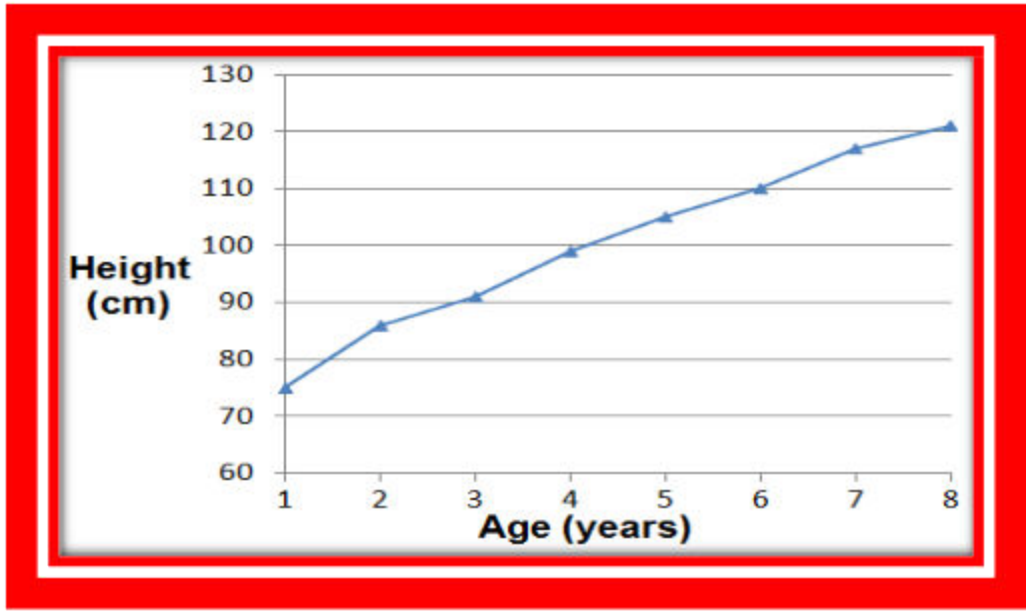
INDEPENDENT VARIABLE



If one variable depends upon or is a consequence of the other variable, it is termed as a dependent variable, and the variable that is antecedent to the dependent variable

IMPORTANT CONCEPTS RELATING TO RESEARCH DESIGN

EXAMPLE



For instance, if we say that height depends upon age, then height is a dependent variable and age is an independent variable.

IMPORTANT CONCEPTS RELATING TO RESEARCH DESIGN

Example for Independent and Dependent Variable:

Age → Height,
Smoking → Cancer
(Height and Cancer are dependent variables whereas Age and Smoking are independent variables).

IMPORTANT CONCEPTS RELATING TO RESEARCH DESIGN

Extraneous variable

Independent variables that are not related to the purpose of the study, but may affect the dependent variable are termed as **extraneous variables.**

Whatever effect is noticed on dependent variable as a result of extraneous variable(s) is technically described as '***experimental error***'.

IMPORTANT CONCEPTS RELATING TO RESEARCH DESIGN

Suppose the researcher wants to test the hypothesis that there is a relationship between student working hours and exam performances. In this case student working hours is an independent variable and exam performances is a dependent variable.

Intelligence may as well affect the exam performances, but since it is not related to the purpose of the study undertaken by the researcher, it will be termed as an extraneous variable.

IMPORTANT CONCEPTS RELATING TO RESEARCH DESIGN

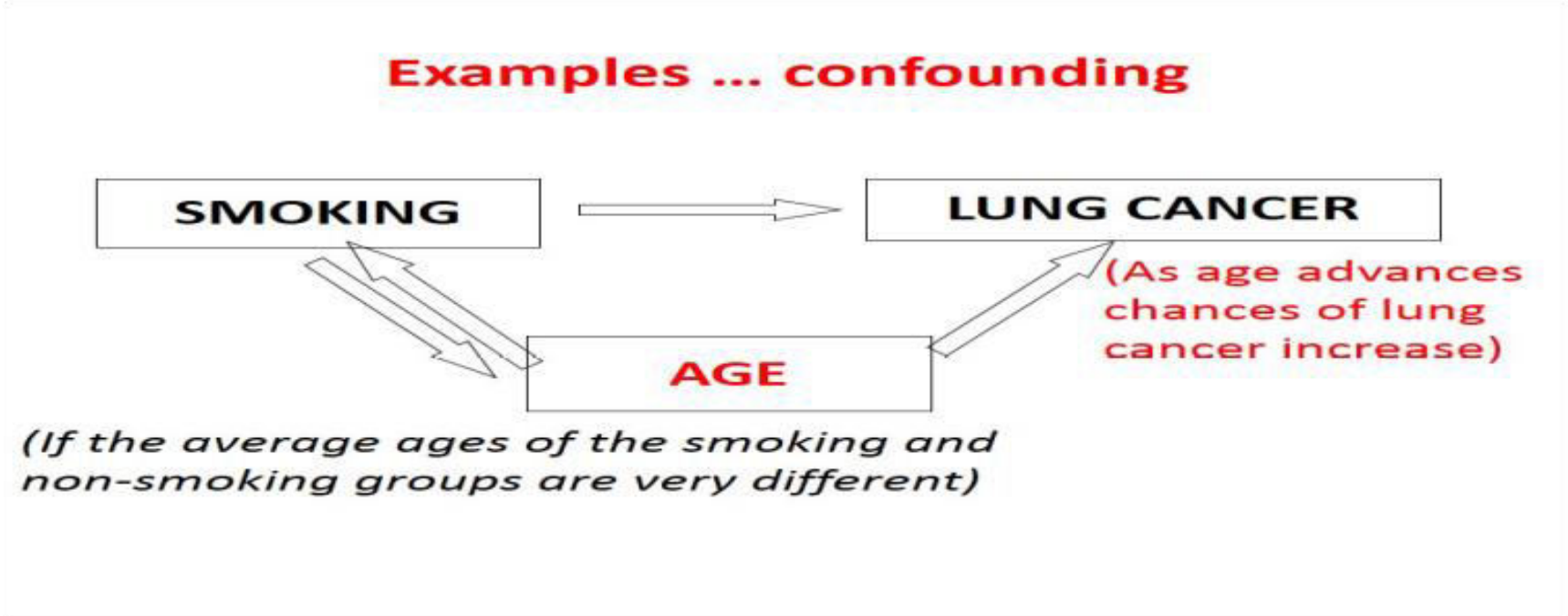
CONTROL :

The technical term **'control'** is used when we design the study minimising the effects of extraneous independent variables.
In experimental researches, the term **'control'** is used to refer to restrain experimental conditions.

IMPORTANT CONCEPTS RELATING TO RESEARCH DESIGN

ConfoundedRelationship

When the dependent variable is not free from the influence of extraneous variable(s), the relationship between the dependent and independent variables is said to be confounded by an extraneous variable(s).



IMPORTANT CONCEPTS RELATING TO RESEARCH DESIGN

Research Hypothesis

- When a prediction or a hypothesised relationship is to be tested by scientific methods, it is termed as ‘research hypothesis’.
- The research hypothesis is a predictive statement that relates an independent variable to a dependent variable.
- Usually, a research hypothesis must contain at least, one independent and one dependent variable.
- For example, “e-Learning enhances teaching learning experience”. Here, the dependent variable is “teaching learning experience”, whereas “e-Learning” is the independent variable.
- *Listening to music lowers blood pressure level.*

IMPORTANT CONCEPTS RELATING TO RESEARCH DESIGN

Experimental and Non-experimental Hypothesis-Testing Research

- Purpose = Test Research Hypothesis:

1) Experimental Hypothesis-testing Research – independent variable is manipulated
2) Non-experimental Hypothesis-testing Research – independent variable NOT manipulated

Example:
Effectiveness of a training programme on the student's performance-level:

- Researcher randomly selects 50 students from a group of students who are to take a course in statistics and then divides them into two groups by randomly assigning 25 to Group A, the usual studies programme, and 25 to Group B, the special studies programme.
- At the end of the course, he administers a test to each group in order to judge the effectiveness of the training programme on the student's performance-level.
- (Exp Hypothesis testing) Why? The type of training programme is manipulated.

IMPORTANT CONCEPTS RELATING TO RESEARCH DESIGN

Experimental and Non-experimental Hypothesis-Testing Research

- Purpose = Test Research Hypothesis:

1) Experimental Hypothesis-testing Research – independent variable is manipulated
2) Non-experimental Hypothesis-testing Research – independent variable NOT manipulated

Example:

For instance, suppose a researcher wants to study whether intelligence affects reading ability for a group.
of students and for this purpose he randomly selects 50 students and tests their intelligence and reading ability by calculating the coefficient of correlation between the two sets of scores. This is an example of non-experimental hypothesis-testing research because herein the independent variable, intelligence, is not manipulated.

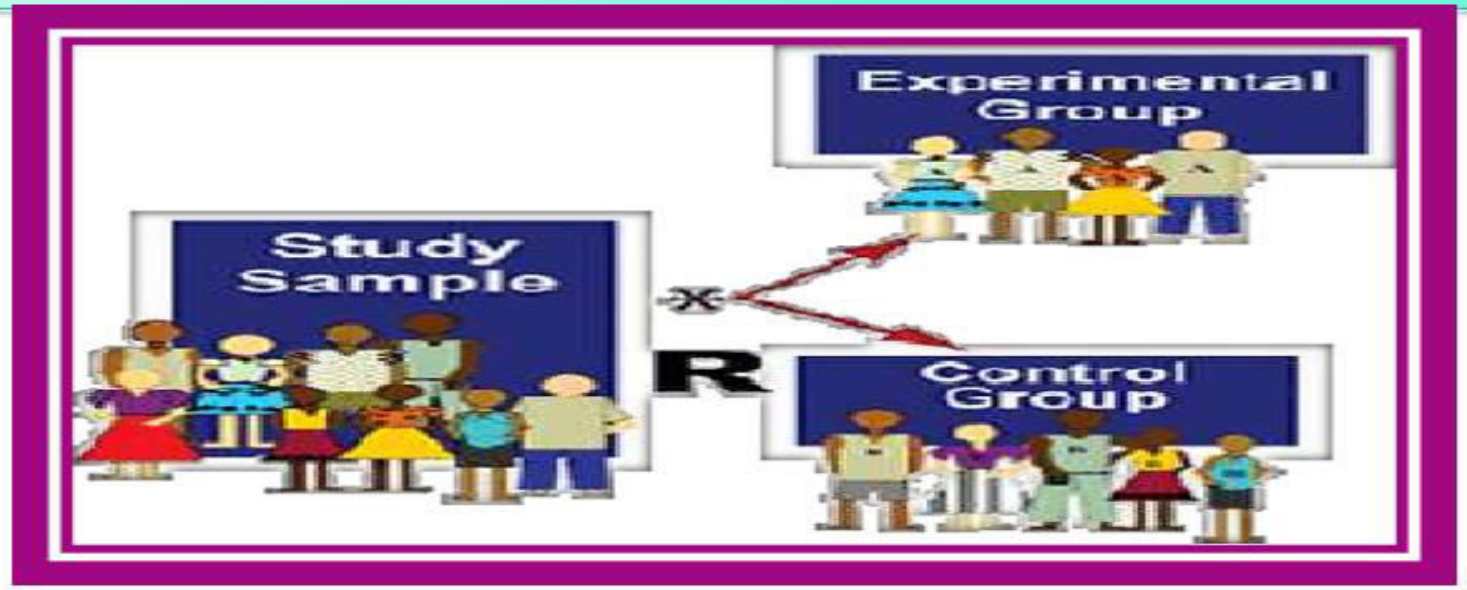
IMPORTANT CONCEPTS RELATING TO RESEARCH DESIGN

- Experimental and control groups
  - In an experimental hypothesis-testing research when a group is exposed to usual conditions, it is termed as 'control group', but when a group is exposed to some novel or special condition, it is termed as 'experimental group'.

Example:
A student is testing to see if plants will grow without sunlight.
Which would be the experimental group and which would be the control group?

IMPORTANT CONCEPTS RELATING TO RESEARCH DESIGN

In the above illustration, the Group A can be called a control group and the Group B an experimental group. If both groups A and B are exposed to special studies programmes, then both groups would be termed 'experimental groups.' It is possible to design studies which include only experimental groups or studies which include both experimental and control groups.



IMPORTANT CONCEPTS RELATING TO RESEARCH DESIGN

Treatment

The different conditions under which experimental and control groups are put are usually referred to as 'treatments'.

In the illustration taken above, the two treatments are the usual studies programme and the special studies programme. Similarly, if we want to determine through an experiment the comparative impact of three varieties of fertilizers on the yield of wheat, in that case the three varieties of fertilizers will be treated as three treatments.

IMPORTANT CONCEPTS RELATING TO RESEARCH DESIGN

IMPORTANT CONCEPTS RELATING TO RESEARCH DESIGN

- Experiment

The process of examining the truth of a statistical hypothesis, relating to some research problem is known as an experiment.

- For example, we can conduct an experiment to examine the usefulness of a newly developed fertilizer. Experiments can be of two types viz., absolute experiment and comparative experiment.
- Absolute experiment – Determining the impact of a fertilizer on the yield of a crop.
- Comparative experiment – Determining the impact of one fertilizer as compared to the impact of some other fertilizer.

IMPORTANT CONCEPTS RELATING TO RESEARCH DESIGN

Experimental unit(s):

The pre-determined plots or the blocks, where different treatments are used, are known as experimental units. Such experimental units must be selected (defined) very carefully.
a physical entity that is the primary unit of interest in a specific research objective.

DIFFERENT RESEARCH DESIGNS

Different research designs can be conveniently described if we categorize them as:

(1) research design in case of exploratory research studies;
(2) research design in case of descriptive and diagnostic research studies, and
(3) research design in case of hypothesis-testing research studies.

DIFFERENT RESEARCH DESIGNS

(1) Research design in case of exploratory research studies

Exploratory research studies are also termed as formulative research studies. The main purpose of such studies is that of formulating a problem for more precise investigation or of developing the working hypotheses from an operational point of view.

The major emphasis in such studies is on the discovery of ideas and insights. The research design appropriate for such studies must be flexible enough to provide opportunity for considering different aspects of a problem under study.

Inbuilt flexibility in research design is needed because the research problem, broadly defined initially, is transformed into one with more precise meaning in exploratory studies, which fact may necessitate changes in the research procedure for gathering relevant data. Generally, the following three methods in the context of research design for such studies are talked about:
(a) the survey of concerning literature; (b) the experience survey and (c) the analysis of 'insight-stimulating' examples.

DIFFERENT RESEARCH DESIGNS

(2) Research design in case of descriptive and diagnostic research studies

Descriptive research studies are those studies which are concerned with describing the characteristics of a particular individual, or of a group, whereas diagnostic research studies determine the frequency with which something occurs or its association with something else.

The studies concerning whether certain variables are associated are examples of diagnostic research studies.

DIFFERENT RESEARCH DESIGNS

Research design in case of descriptive and diagnostic research studies

The design in such studies must be rigid and not flexible and must focus attention on the following:

(a) Formulating the objective of the study (what the study is about and why is it being made?)

(b) Designing the methods of data collection (what techniques of gathering data will be adopted?)

(c) Selecting the sample (how much material will be needed?)

(d) Collecting the data (where can the required data be found and with what time period should the data be related?)

(e) Processing and analysing the data.

(f) Reporting the findings

DIFFERENT RESEARCH DESIGNS

Research design in case of descriptive and diagnostic research studies

Research Design	*Type of study*	
	Exploratory of Formulative	*Descriptive/Diagnostic*
Overall design	Flexible design (design must provide opportunity for considering different aspects of the problem)	Rigid design (design must make enough provision for protection against bias and must maximise reliability)
(i) Sampling design	Non-probability sampling design (purposive or judgement sampling)	Probability sampling design (random sampling)
(ii) Statistical design	No pre-planned design for analysis	Pre-planned design for analysis
(iii) Observational design	Unstructured instruments for collection of data	Structured or well thought out instruments for collection of data
(iv) Operational design	No fixed decisions about the operational procedures	Advanced decisions about operational procedures.

DIFFERENT RESEARCH DESIGNS

(3) Research design in case of hypothesis-testing research studies:

Hypothesis-testing research studies are those where the researcher tests the hypotheses of causal relationships between variables.

Such studies require procedures that will not only reduce bias and increase reliability, but will permit drawing inferences about causality.

Usually experiments meet this requirement. Hence, when we talk of research design in such studies, we often mean the design of experiments.

Basic Principles of Experimental Design

Professor Fisher has enumerated three principles of experimental designs:

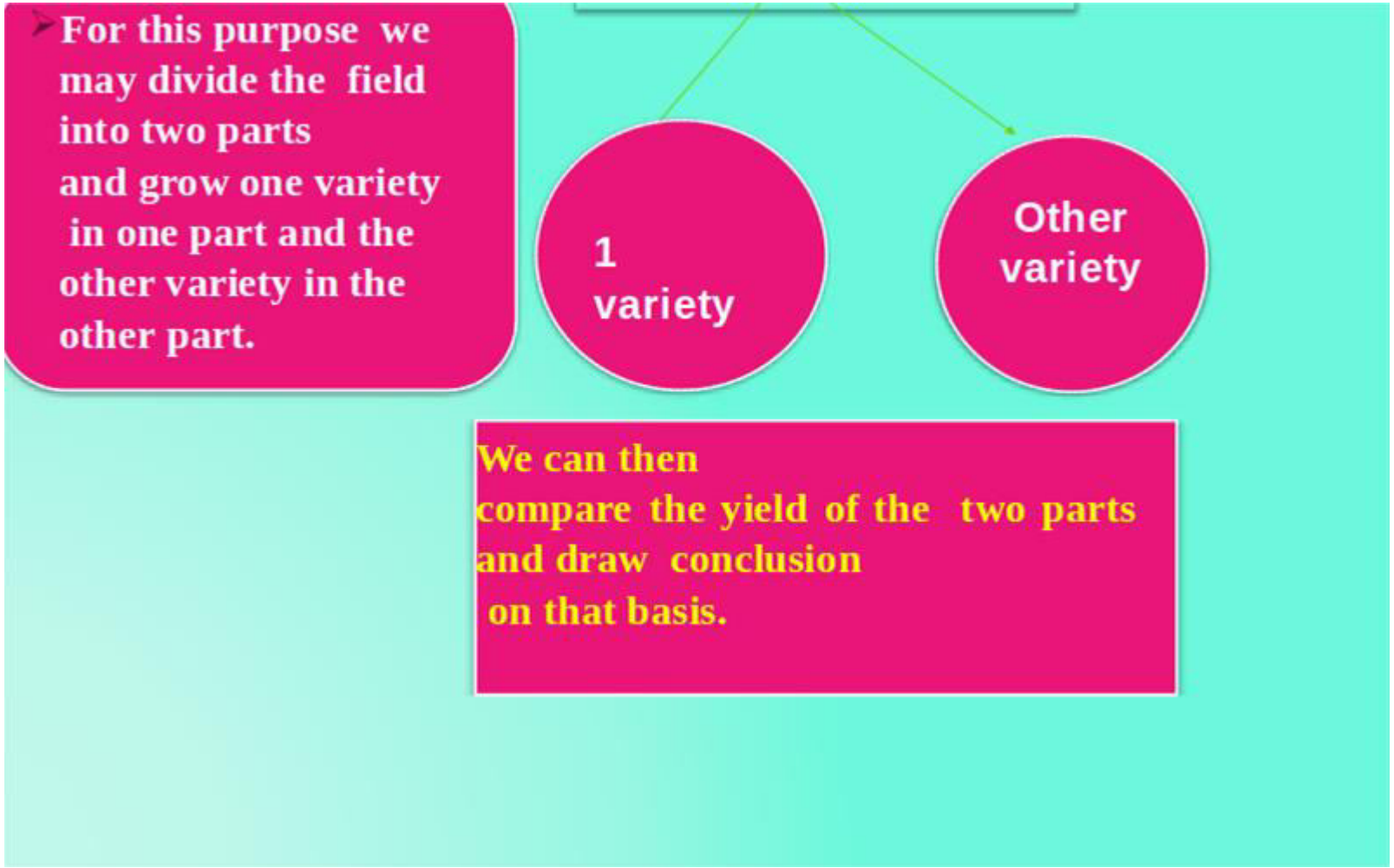
1. Principle of Replication
2. Principle of Randomization
3. Principle of Local Control

Basic Principles of Experimental Design

[1] PRINCIPLE OF REPLICATION

According to the *Principle of Replication*, the experiment should be repeated more than once. Thus, each treatment is applied in many experimental units instead of one. By doing so the statistical accuracy of the experiments is increased

Basic Principles of Experimental Design



Basic Principles of Experimental Design

PRINCIPLE OF RANDOMIZATION

The *Principle of Randomization* provides protection, when we conduct an experiment, against the effect of extraneous factors by randomization. In other words, this principle indicates that we should design or plan the experiment in such a way that the variations caused by extraneous factors can all be combined under the general heading of "chance."

Basic Principles of Experimental Design

- **Suppose** , if we grow one variety of rice

two varieties of rice

- For this purpose we may divide the field into two parts and grow one variety in one part and the other variety in the other part.

1 variety

Other variety

then it is just possible that the soil fertility may be different in the first half in comparison to the other half. If this is so, our results would not be realistic

Basic Principles of Experimental Design

[2] PRINCIPLE OF LOCAL CONTROL

The *Principle of Local Control* is another important principle of experimental designs. Under it the extraneous factor, the known source of variability, is made to vary deliberately over as wide a range as necessary and this needs to be done in such a way that the variability it causes can be measured and hence eliminated from the experimental error

Basic Principles of Experimental Design

- **Suppose** , if we perform a two-way analysis of variance,

in which the total variability of the data is divided into three components attributed to treatments

varieties of rice in our case

the extraneous factor (soil fertility in our case

experimental error

In brief, through the principle of local control we can eliminate the variability due to extraneous factors from the experimental error.

Important Experimental Design

Experimental design refers to the framework or structure of an experiment and as such there are several experimental designs.

We can classify experimental designs into two broad categories :-
Informal Experimental Designs:

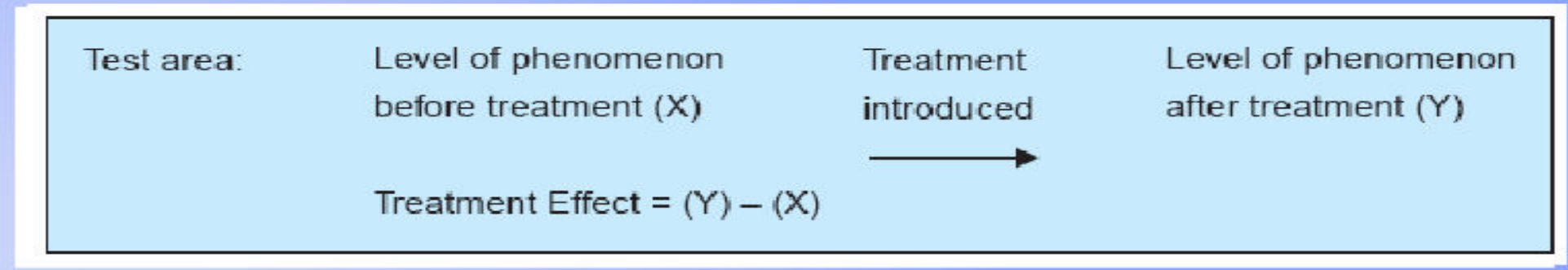
1. Before-and-after without control design.
2. After-only with control design.
3. Before-and-after with control design

Formal Experimental Designs:

1. Completely Randomized Design (C. R. Design)
2. Randomized Block Design (R. B. Design)
3. Latin Square Design (L. S. Design)
4. Factorial Designs

Important Experimental Designs

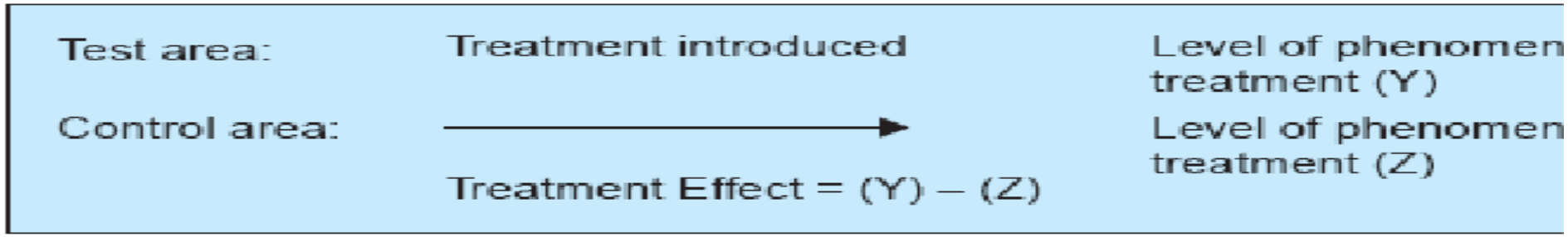
- The treatment is then introduced and then the dependent variable is measured again .
  - The effect of the treatment
    - the level of the phenomenon after the treatment - the lvd of the phenomenon before the treatment .
      - The design can be represented thus:



Important Experimental Design

2. After-only with control design

- **Two groups or areas (test area and control area) ar**
  and the treatment is introduced into the test area



Important Experimental Design

3.Before and after with Control design

In this design two areas are selected and the dependent variable is measured in both the areas for an identical time-period before the treatment. The treatment is then introduced into the test area only, and the dependent variable is measured in both for an identical time-period after the introduction of the treatment

The design can be represented thus:

	Time Period I		Time Period II
Test area:	Level of phenomenon before treatment (X)	Treatment introduced →	Level of phenomenon after treatment (Y)
Control area:	Level of phenomenon without treatment (A)		Level of phenomenon without treatment (Z)

Treatment Effect = (Y – X) – (Z – A)

Important Experimental Design

FORMAL EXPERIMENTAL DESIGNS

[A]Completely randomized design (C.R. design)

Involves only two principles i.e the principle of replication and the principle of randomization of experimental designs.

Important Experimental Design

Completely randomized design (C.R. design):

- Involves only two principles viz., the principle of replication and the principle of randomization.
- It is the simplest possible design and its procedure of analysis is also easier.
- The essential characteristic of the design is that subjects are randomly assigned to experimental treatments (or vice-versa).

Example

10 subjects and if we wish to test 5 under treatment A and 5 under treatment B, the randomization process gives every possible group of 5 subjects selected from a set of 10 an equal opportunity of being assigned to treatment A and treatment B.

- It provides maximum number of degrees of freedom to the error.
- Such a design is generally used when experimental areas happen to be homogeneous.

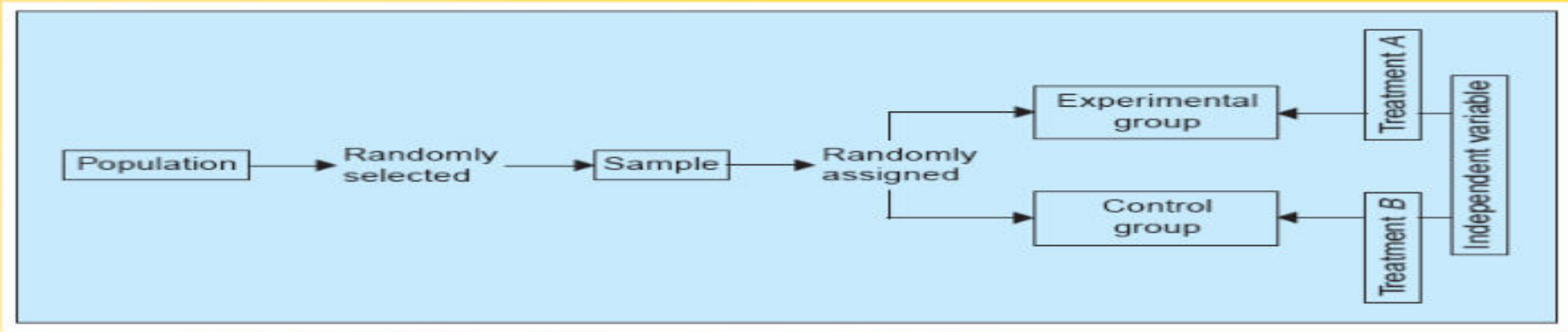
Important Experimental Design
Completely Randomized Design
(1) Two Group Simple Randomized Design

We can present a brief description of the two forms

of such a design as :-

[1] Two-group simple randomized design

first of all the population is defined and then from the population a sample is selected randomly . After being selected randomly from the population, be randomly assigned to the experimental and control groups (Such random assignment of items to two groups is technically described as principle of randomization)

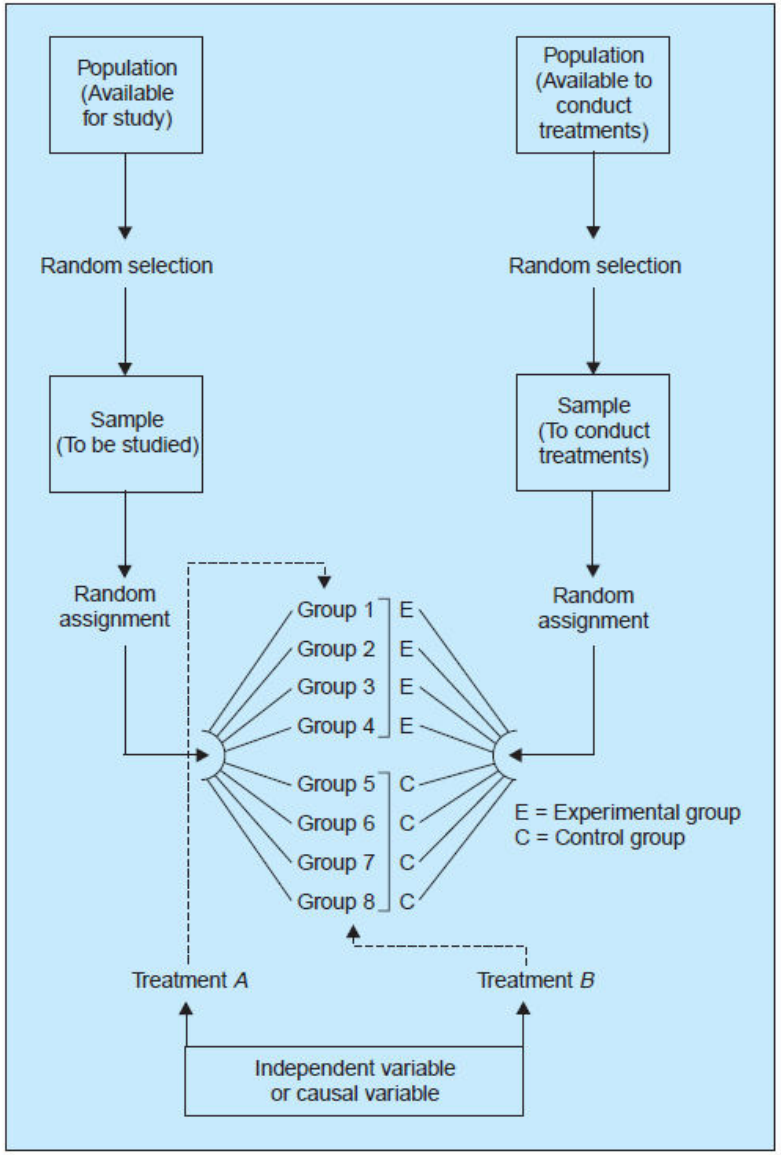


Important Experimental Design
[A] Completely Randomized Design

(2) Random replications design

- The limitation of the two-group randomized design is usually eliminated within the random replications design. In the illustration just cited above, the *teacher differences* on the dependent variable were ignored, i.e., the extraneous variable was not controlled. But in a random replications design, the effect of such differences are minimised (or reduced) by providing a number of repetitions for each treatment. Each repetition Is technically called a 'replication'.

Important Experimental Design
Completely Randomized Design: Random Replications Design



Important Experimental Design
[B] Randomized Block Design

[B] Randomized block design (R.B. design)

- In the R.B. design the principle of *local control* can be applied along with the other two principles of experimental designs.
- In the R.B. design, subjects are first divided into groups, known as blocks. In general, blocks are the levels at which we hold the extraneous factor fixed, so that its contribution to the total variability of data can be measured. The main feature of the R.B. design is that in this each treatment appears the same number of times in each block. The R.B. design is analysed by the two-way analysis of variance (two-way ANOVA)* technique.

Important Experimental Design
[B] Randomized Block Design

Suppose four different forms of a standardised test in statistics were given to each of five students (selected one from each of the five I.Q. blocks) and following are the scores which they obtained.

	Very low I.Q.	Low I.Q.	Average I.Q.	High I.Q.	Very high I.Q.
	Student A	Student B	Student C	Student D	Student E
Form 1	82	67	57	71	73
Form 2	90	68	54	70	81
Form 3	86	73	51	69	84
Form 4	93	77	60	65	71

If each student separately randomized the order in which he or she took the four tests (by using random numbers or some similar device), we refer to the design of this experiment as a R.B. design. The purpose of this randomization is to take care of such possible extraneous factors (say as fatigue) or perhaps the experience gained from repeatedly taking the test.

Important Experimental Design
[C] Latin Square Design

[C] Latin square design (L.S. design)

- **It is an experimental design very frequently used in agricultural research.**

Seeds differences	FERTILITY LEVEL I	II	III	IV	V
X_1	A	B	C	D	E
X_2	B	C	D	E	A
X_3	C	D	E	A	B
X_4	D	E	A	B	C
X_5	E	A	B	C	D

Important Experimental Design
[D] Factorial Designs

[D] FACTORIAL DESIGNS

- Factorial designs are used in experiments where the effects of varying more than one factor are to be determined. They are specially important in several economic and social phenomena where usually a large number of factors affect a particular problem.
- Factorial designs can be of two types:
  i) simple factorial designs
  ii) complex factorial designs

Important Experimental Design
[D] Factorial Designs

Simple Factorial Design

- In case of simple factorial designs, we consider the effects of varying two factors on the dependent variable, but when an experiment is done with more than two factors, we use complex factorial designs.
- Simple factorial design is also termed as a 'two-factor-factorial design', whereas complex factorial design is known as 'multifactor-factorial design.'

Important Experimental Design
[D] Factorial Designs

Simple Factorial Design

2 × 2 SIMPLE FACTORIAL DESIGN

Control variables	Experimental Variable: Treatment A	Experimental Variable: Treatment B
Level I	Cell 1	Cell 3
Level II	Cell 2	Cell 4

Important Experimental Design
[D] Factorial Designs
Complex Factorial Design

- Experiments with more than two factors at a time involve the use of complex factorial designs. A design which considers three or more independent variables simultaneously is called a *complex factorial design.*
- In case of three factors with one experimental variable having two treatments and two control variables, each one of which having two levels, the design used will be termed 2 × 2 × 2 complex factorial design which will contain a total of eight cells.

Important Experimental Design
[D] Factorial Designs
Complex Factorial Design

2 × 2 × 2 COMPLEX FACTORIAL DESIGN

		Experimental Variable			
		Treatment A		Treatment B	
		Control Variable 2 Level I	Control Variable 2 Level II	Control Variable 2 Level I	Control Variable 2 Level II
Control Variable 1	Level I	Cell 1	Cell 3	Cell 5	Cell 7
	Level II	Cell 2	Cell 4	Cell 6	Cell 8

Important Experimental Design

CONCLUSION

There are several research designs and the researcher must decide in advance of collection and analysis of data as to which design would prove to be more appropriate for his research project. He must give due weight to various points such as the type of universe and its nature, the objective of his study, the resource list or the sampling frame, desired standard of accuracy and the like when taking a decision in respect of the design for his research project.

PES UNIVERSITY ONLINE

RESEARCH METHODOLOGY

Research Design

RESEARCH METHODOLOGY

Sampling Design

Priya Badarinath & Raghu B. A & Roopa Ravish

Department of Computer Science and Engineering

RESEARCH METHODOLOGY

Ch. 4: Sampling Design

Department of Computer Science and Engineering

CONTENTS

INTRODUCTION

- Items in field of inquiry – Universe / Population
- Enumeration of all items in population = Census Survey.
  - When all objects are covered -> highest accuracy
  - Possible? Cost, Bias?
- Census is impossible in situation where population is infinite.
- Select only few items = Sample
- Procedure for selection = Sampling Technique.

- A **sample** is “a smaller (but hopefully representative) collection of units from a population used to determine truths about that population” (Field, 2005)

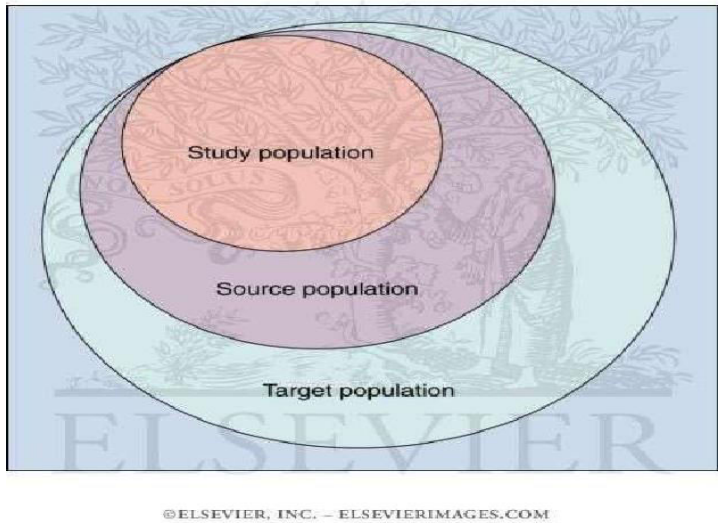
- The **sampling frame** A list of all elements or other units containing the elements in a population.

POPULATION

The larger group from which individuals are selected to participate in a study

TARGET POPULATION

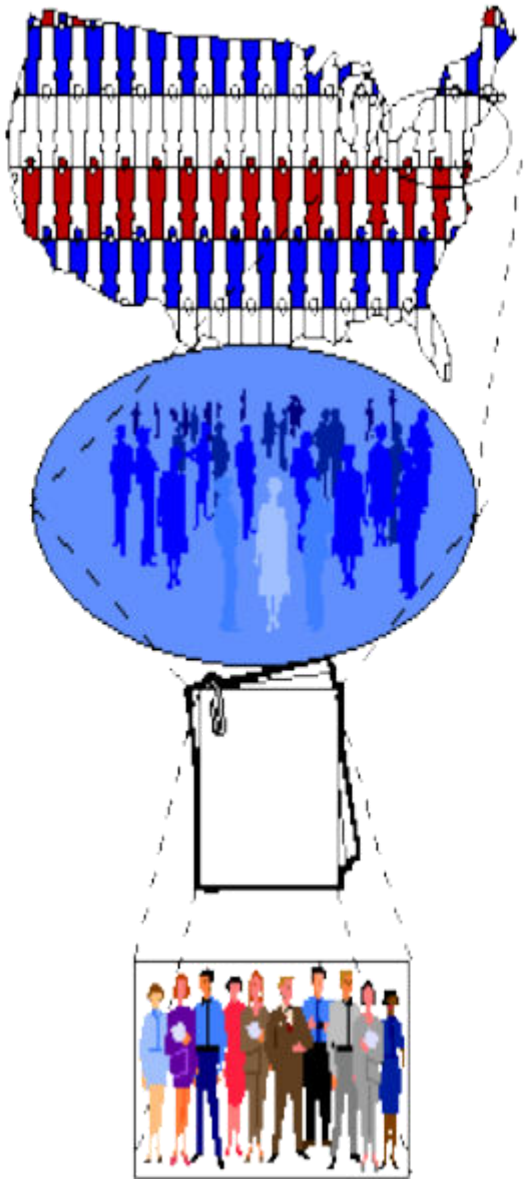
A set of elements larger than or different from the population sampled and to which the researcher would like to generalize study findings.



POPULATION TO SAMPLE

Who do you want to generalize to?

The Theoretical Population



What population can you get access to?

The Study Population

How can you get access to them?

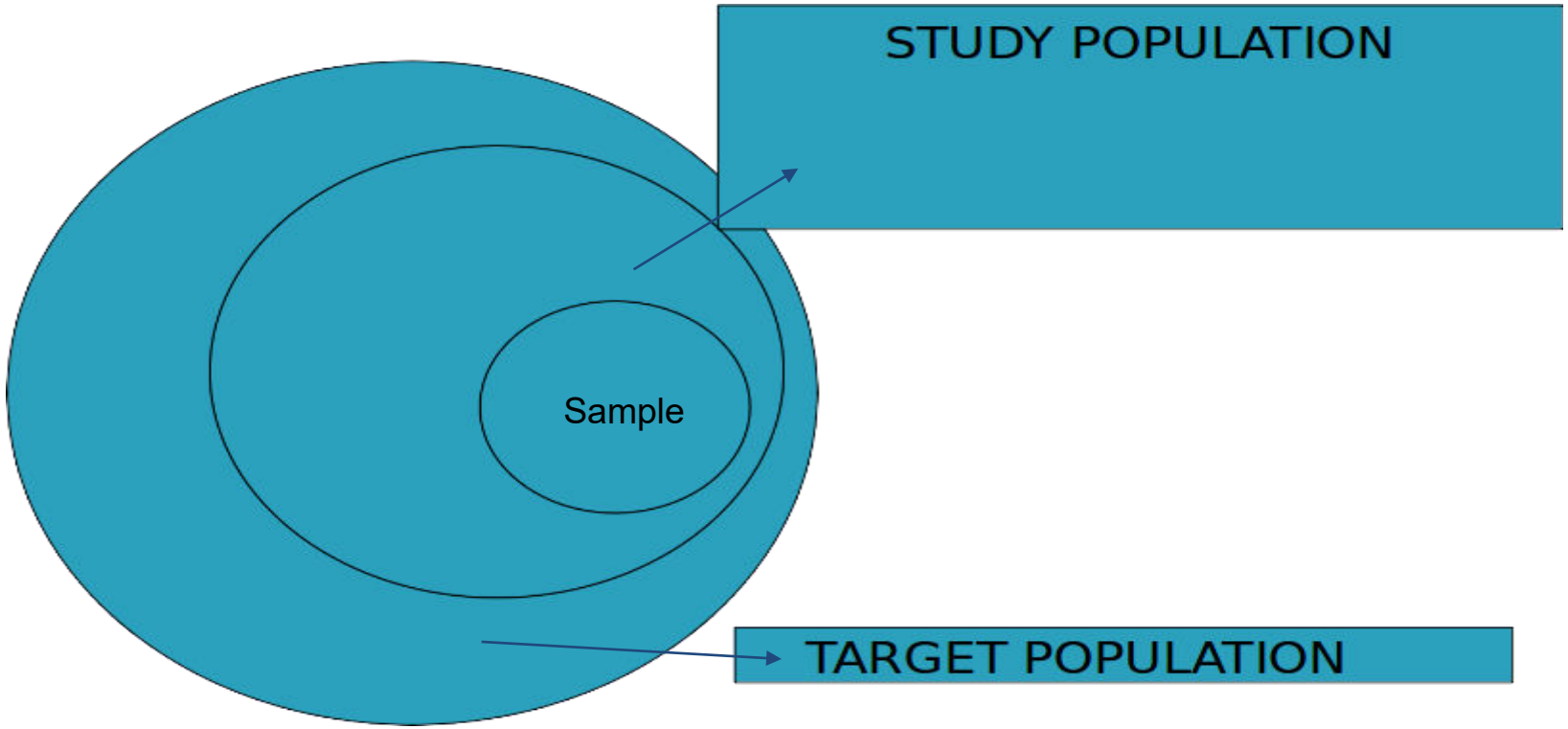
The Sampling Frame

Who is in your study?

The Sample

SAMPLING

Definition: "*The process of selecting a number of individuals for a study in such a way that the individuals represent the larger group from which they were selected.*"



SAMPLE DESIGN

- Plan for obtaining sample from a population
- Technique / procedure a researcher will adopt in selecting sample.
- Sample design is determined before data are collected.
- There are many sample designs from which a researcher can choose.
- Some designs are relatively more precise and easier to apply than others.
- Researcher must select/prepare a sample design which should be reliable and appropriate for his research study.

MAIN STEPS OF SAMPLING

1) Objective
2) Population
3) Sampling Unit and Frame
4) Size of Sample
5) Parameter of Interest
6) Data Collection
7) Non respondents
8) Selection of proper sampling design
9) Organizing field work
10) Pilot Survey
11) Budgetary Constraints

MAIN STEPS OF SAMPLING

Objective :

- Define the objective of study. It should be in proportion with money, manpower and time.

Population (or Universe):

- Should be clearly defined.

Sampling Units and Frames:

- Sampling unit – for study
  - geographic unit – state, district, state or
  - construction unit – flat, house, or
  - social unit – house, flat

Source List – Sampling Frame

MAIN STEPS OF SAMPLING

Size of Sample

- Number of units to be selected.
- Size should be neither too large nor too small
- Optimal size –> efficiency, flexibility, reliability.

Parameters of Interest

- Statistical constants of population – parameters.
- Ex - mean population, population proportion

MAIN STEPS OF SAMPLING

Data Collection:

- Only relevant information to be collected.
- Objective to be made clear

Non respondents

- Because of practical difficulties, data may not be collected for all sampled units.
- Non responses tend to change results

MAIN STEPS OF SAMPLING

Selection of proper Sampling Design :

- Must decide technique in selecting the items for sample.
- Must yield less error

Organizing field work:

- Success depends on reliable field work.
- There should be efficient supervisory staff and trained personnel for field work.

MAIN STEPS OF SAMPLING

Pilot Survey

- Try research on small scale before going to field.

Budgetary Constraints

- Cost consideration, from practical viewpoint.
- Have major impact upon decision relating to not only size but also types of sample.

Organizing field work

- Success depends on reliable field work

CRITERIA OF SELECTING A SAMPLING PROCEDURE

Two costs are involved in a sampling analysis viz., the cost of collecting the data and the cost of an incorrect inference resulting from the data. Researcher must keep in view the two causes of incorrect inferences viz., systematic bias and sampling error.

CRITERIA OF SELECTING A SAMPLING PROCEDURE

1. **Inappropriate sampling frame:** If the sampling frame is inappropriate i.e., a biased representation of the universe, it will result in a systematic bias.

2. Defective measuring device: If the measuring device is constantly in error, it will result in systematic bias. In survey work, systematic bias can result if the questionnaire or the interviewer is biased. Similarly, if the physical measuring device is defective there will be systematic bias in the data collected through such a measuring device.

3. Non-respondents: If we are unable to sample all the individuals initially included in the sample, there may arise a systematic bias. The reason is that in such a situation the likelihood of establishing contact or receiving a response from an individual is often correlated with the measure of what is to be estimated.

CRITERIA OF SELECTING A SAMPLING PROCEDURE

4. Indeterminancy principle: Sometimes we find that individuals act differently when kept under observation than what they do when kept in non-observed situations.

For instance, if workers are aware that somebody is observing them in course of a work study on the basis of which the average length of time to complete a task will be determined and accordingly the quota will be set for piece work, they generally tend to work slowly in comparison to the speed with which they work if kept unobserved. Thus, the indeterminancy principle may also be a cause of a systematic bias.

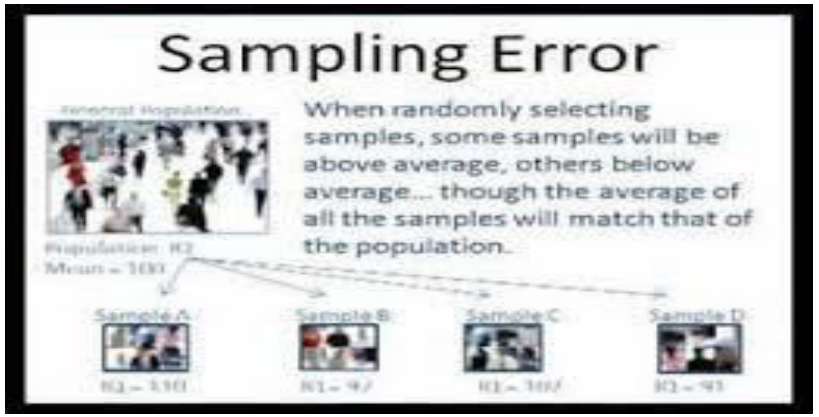
CRITERIA OF SELECTING A SAMPLING PROCEDURE

5. Natural bias in the reporting of data:

Natural bias of respondents in the reporting of data is often the cause of a systematic bias in many inquiries. There is usually a downward bias in the income data collected by government taxation department, whereas we find an upward bias in the income data collected by some social organisation. People in general understate their incomes if asked about it for tax purposes, but they overstate the same if asked for social status or their affluence. Generally in psychological surveys, people tend to give what they think is the ‘correct’ answer rather than revealing their true feelings.

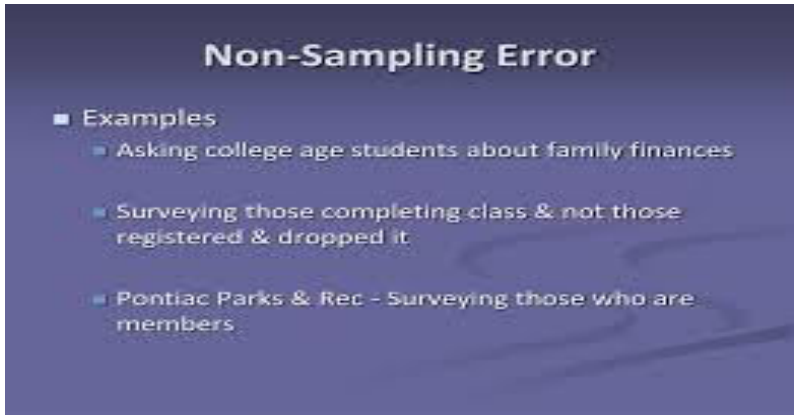
SAMPLING ERROR

- Only a part of population has been used for estimation and to draw inference.
- Sampling errors are absent in census survey
- Can be measured for a given size and sample design = ***precision*** of sampling plan.
- If we increase the sample size, precision can be improved.
- Also increases cost and systematic bias



NON- SAMPLING ERROR

- Arises at the stage of collection and preparation of data.
- Present in both sample & census survey
- Can be reduced by defining the sampling unit, frame and population correctly.



CHARACTERISTICS OF A GOOD SAMPLE DESIGN

(a) Sample design must result in a truly representative sample.

(b) Sample design must be such which results in a small sampling error.

(c) Sample design must be viable in the context of funds available for the research study.

(d) Sample design must be such so that systematic bias can be controlled in a better way.

(e) Sample should be such that the results of the sample study can be applied, in general, for the universe with a reasonable level of confidence.

SAMPLING TECHNIQUES

PES UNIVERSITY ONLINE

TYPES OF SAMPLING DESIGN

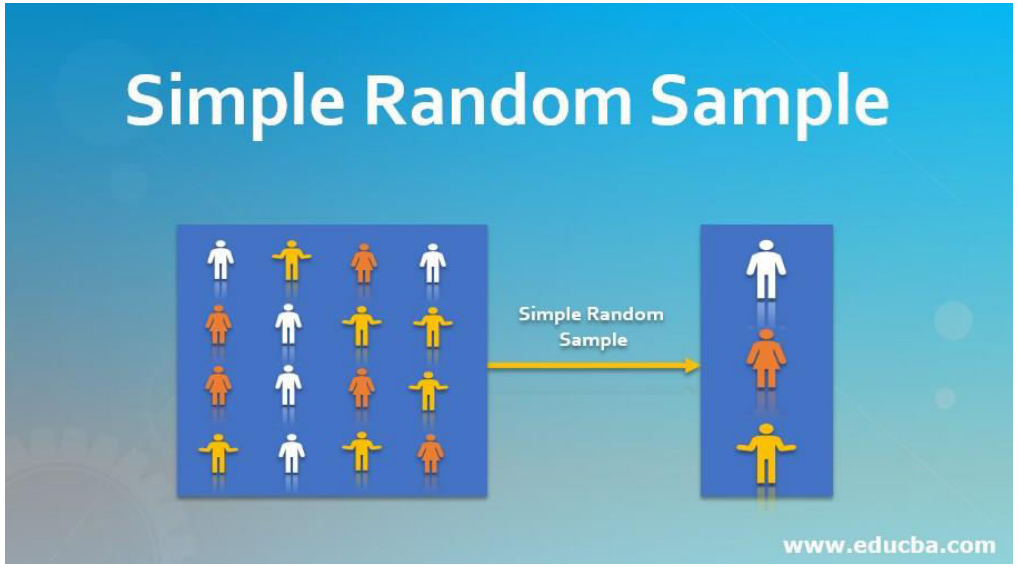
CHART SHOWING BASIC SAMPLING DESIGNS

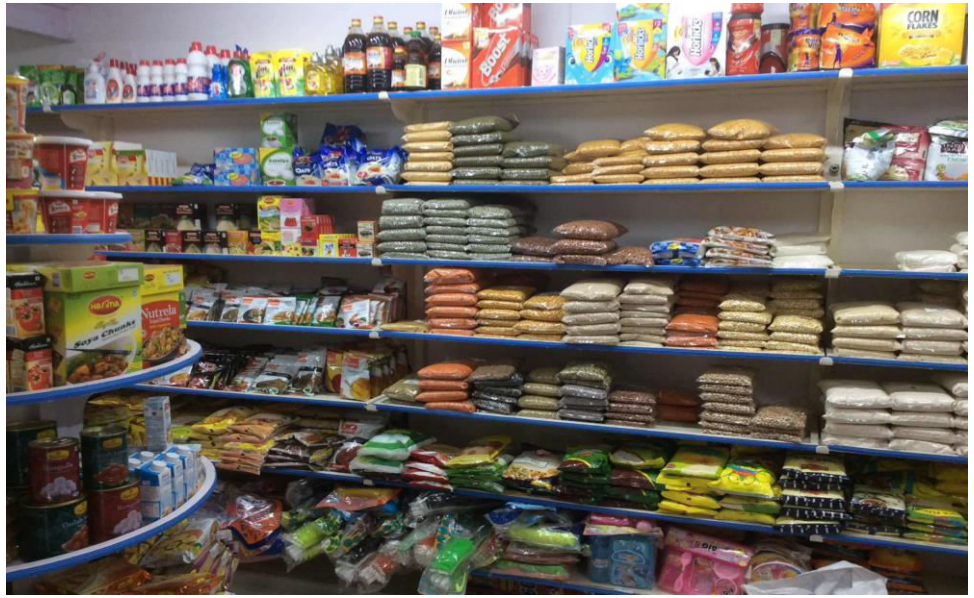
Element selection technique	Representation basis: Probability sampling	Representation basis: Non-probability sampling
Unrestricted sampling	Simple random sampling	Haphazard sampling or convenience sampling
Restricted sampling	Complex random sampling (such as cluster sampling, systematic sampling, stratified sampling etc.)	Purposive sampling (such as quota sampling, judgement sampling)

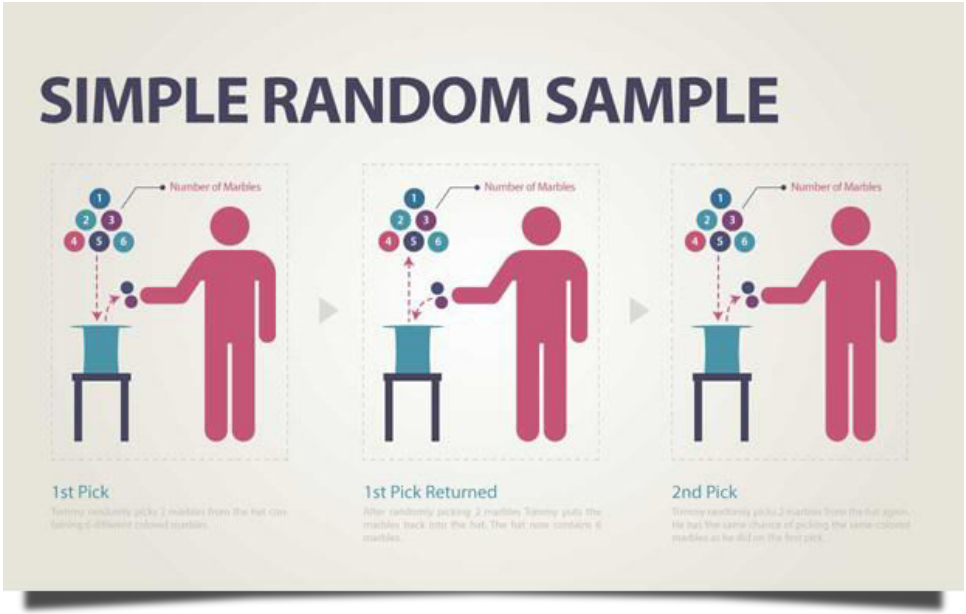
SIMPLE RANDOM SAMPLING

- It is the purest form of probability sampling.
- Each member of the finite population has an equal probability and known chance of being selected.
- Sampling without replacement
- When there are very large populations?
  - it is often difficult
  - impossible to identify every member of the population
  - The pool of available subjects becomes biased.
- Examples :
  - Lottery or Fishbowl
  - Survey of insect population living in woodland
  - Trees no. 1 to 1200. 10 trees are chosen at random
- Random number tables available [by Tippett, Yates, Fisher]

SIMPLE RANDOM SAMPLING





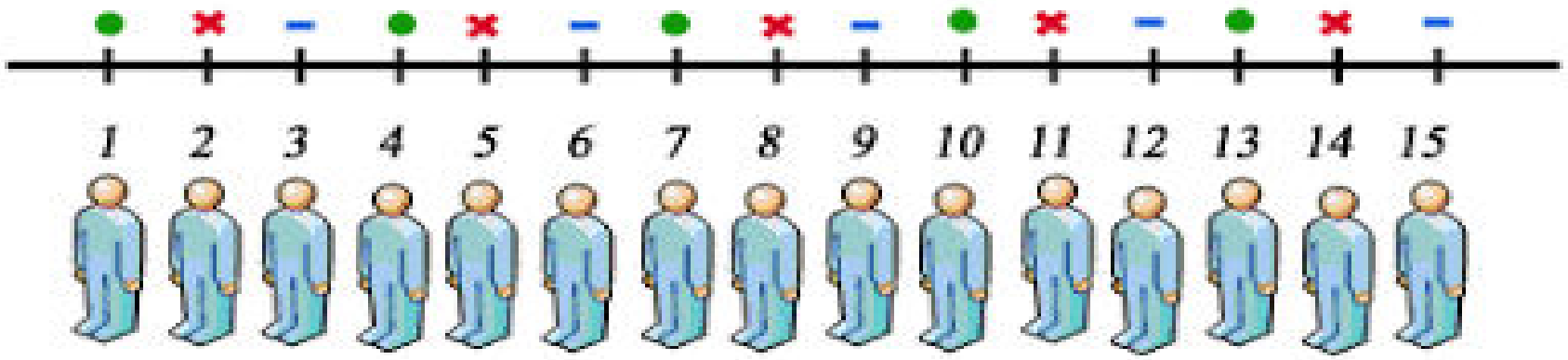


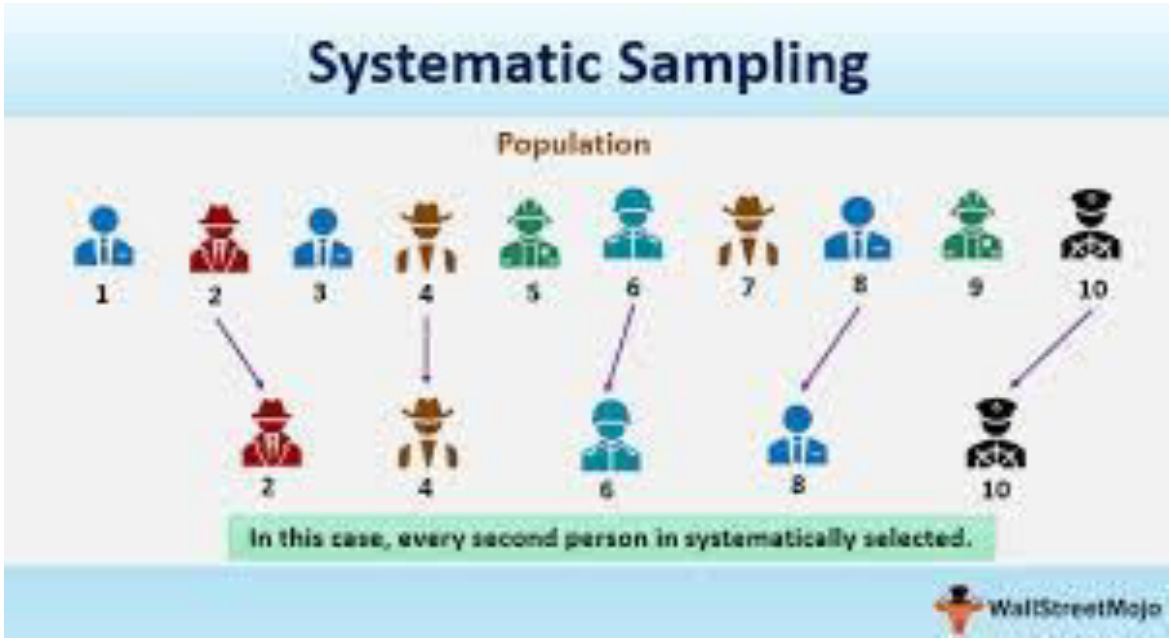
COMPLEX RANDOM SAMPLING DESIGNS

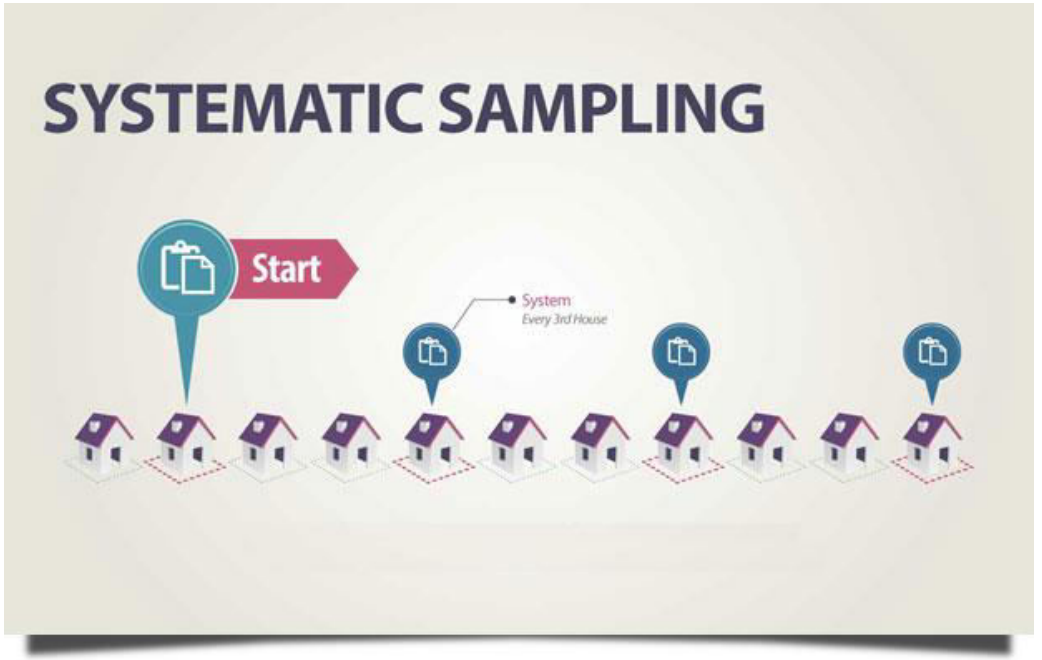
SYSTEMATIC SAMPLING

- It is often used instead of random sampling.
- It is also called an Nth name selection technique.
- After the required *sample size* has been calculated, every Nth record is selected from a list of population members.
- Randomness introduced in picking the start point.
- As long as the list *does not contain any hidden order*, this sampling method is as good as the random sampling method.
- It's only advantage over the random sampling technique is simplicity, easy, cheap. And applicable for large populations.
- Systematic sampling is frequently used to select a specified number of records from a computer file.

SYSTEMATIC SAMPLING







SYSTEMATIC SAMPLING

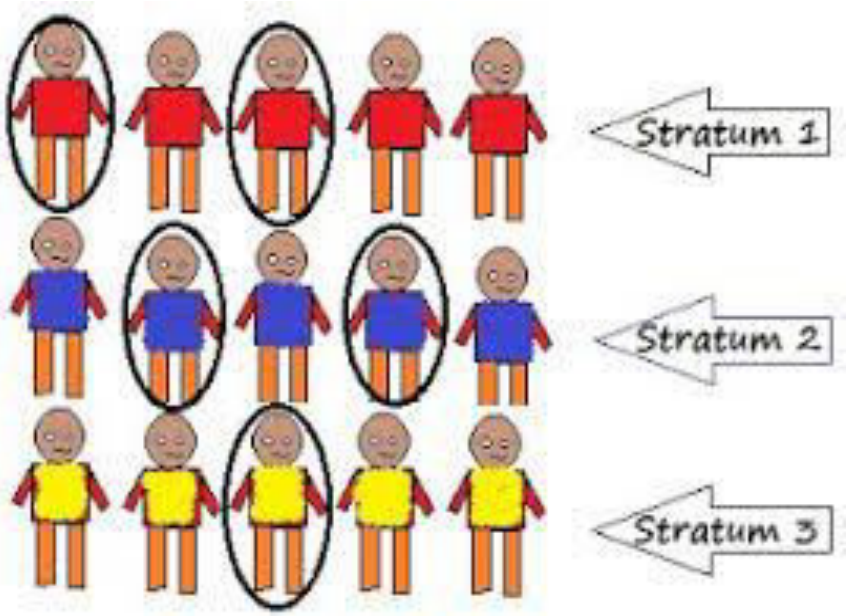
Example: to select a sample of 25 dorm rooms in your college hotel, make a list of all the room numbers in the dorm.

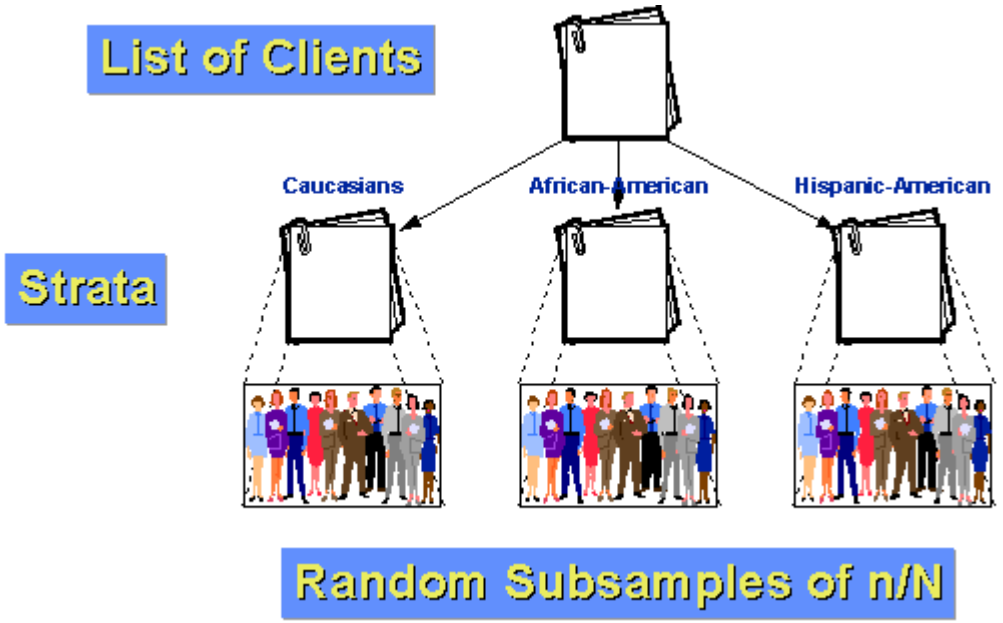
- For example if there are 100 rooms.
- divide the total number of rooms (100) by the number of rooms you want in the sample (25). The answer is 4.
- This means that you are going to *select every fourth* dorm room from the list.
- First of all, we have to determine the random starting point.
- This step can be done by picking any point on the table of room numbers, and read across or down until you come to a number between 1 and 4. This is your random starting point.
- For instance, your random starting point is "3". This means you select dorm room #3 as your first room, and then every fourth room down the list (3, 7, 11, 15, 19, etc.) until you have 25 rooms selected.

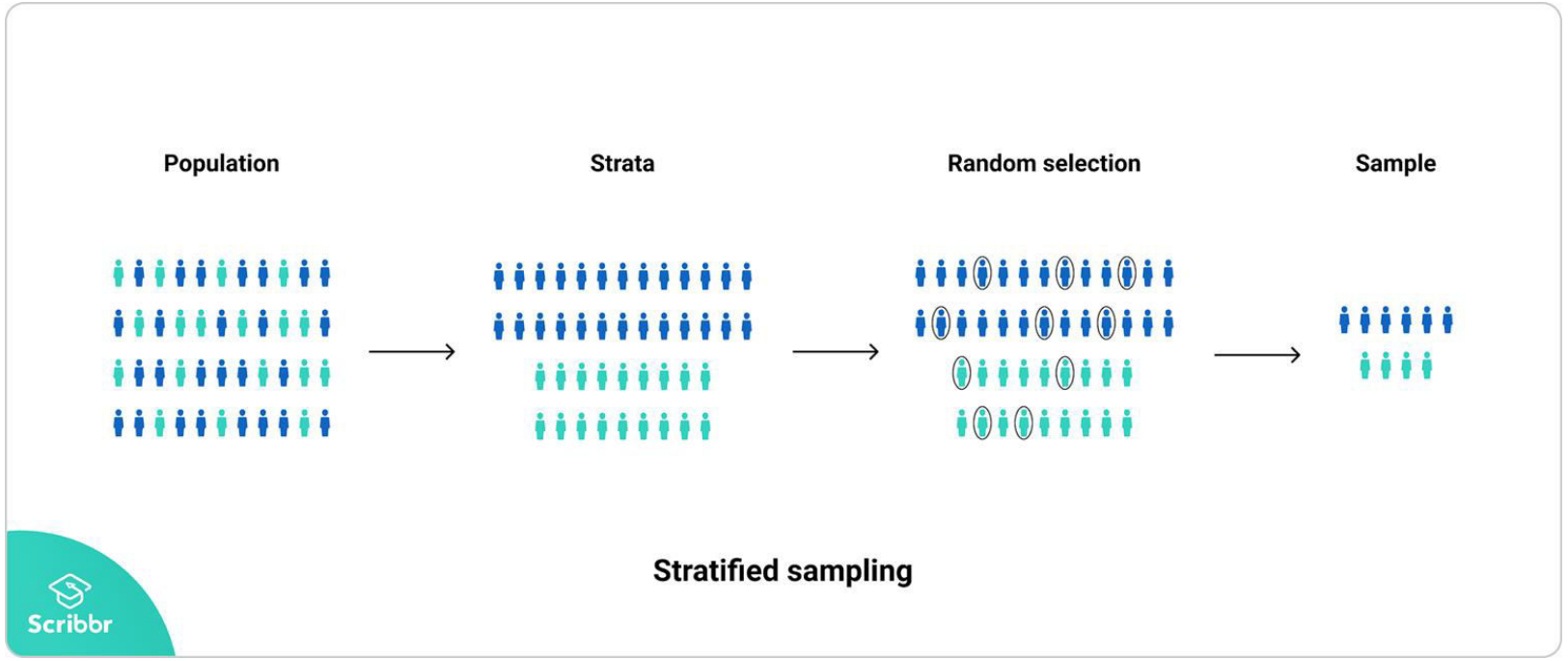
STRATIFIED SAMPLING

- Chosen when sample to be drawn doesn't constitute homogeneous group.
- Population is divided into sub-populations that are individually homogeneous - *strata*
- Stratified Sampling is possible when it makes sense to partition the population into groups based on a factor that may influence the variable that is being measured.
- These groups are then called *strata*. Based on one or more characteristic.
- An individual group is called a stratum. With stratified sampling one should:
    - partition the population into groups (strata)
    - obtain a simple random sample from each group (stratum)
    - collect data on each sampling unit that was randomly sampled from each group (stratum)
    - Called Stratified random sampling
- Stratified sampling works best when a heterogeneous population is split into fairly homogeneous groups.

STRATIFIED SAMPLING







STRATIFIED SAMPLING

	Example 1	Example 2	Example 3
Population	All people in US	All PSU intercollegiate athletes	All elementary students in the local school district
Groups (Strata)	4 Time Zones in the U.S. (Eastern,Central, Mountain,Pacific)	26 PSU intercollegiate teams	11 different elementary schools in the local school district
Obtain a Simple Random Sample	500 people from each of the 4 time zones	5 athletes from each of the 26 PSU teams	20 students from each of the 11 elementary schools
Sample	$4 \times 500 = 2000$ selected people	$26 \times 5 = 130$ selected athletes	$11 \times 20 = 220$ selected students

CLUSTER SAMPLING

- Used when population is divided into groups or clusters
- Samples are selected from groups rather than individuals which is employed into large scale survey.

Advantages:

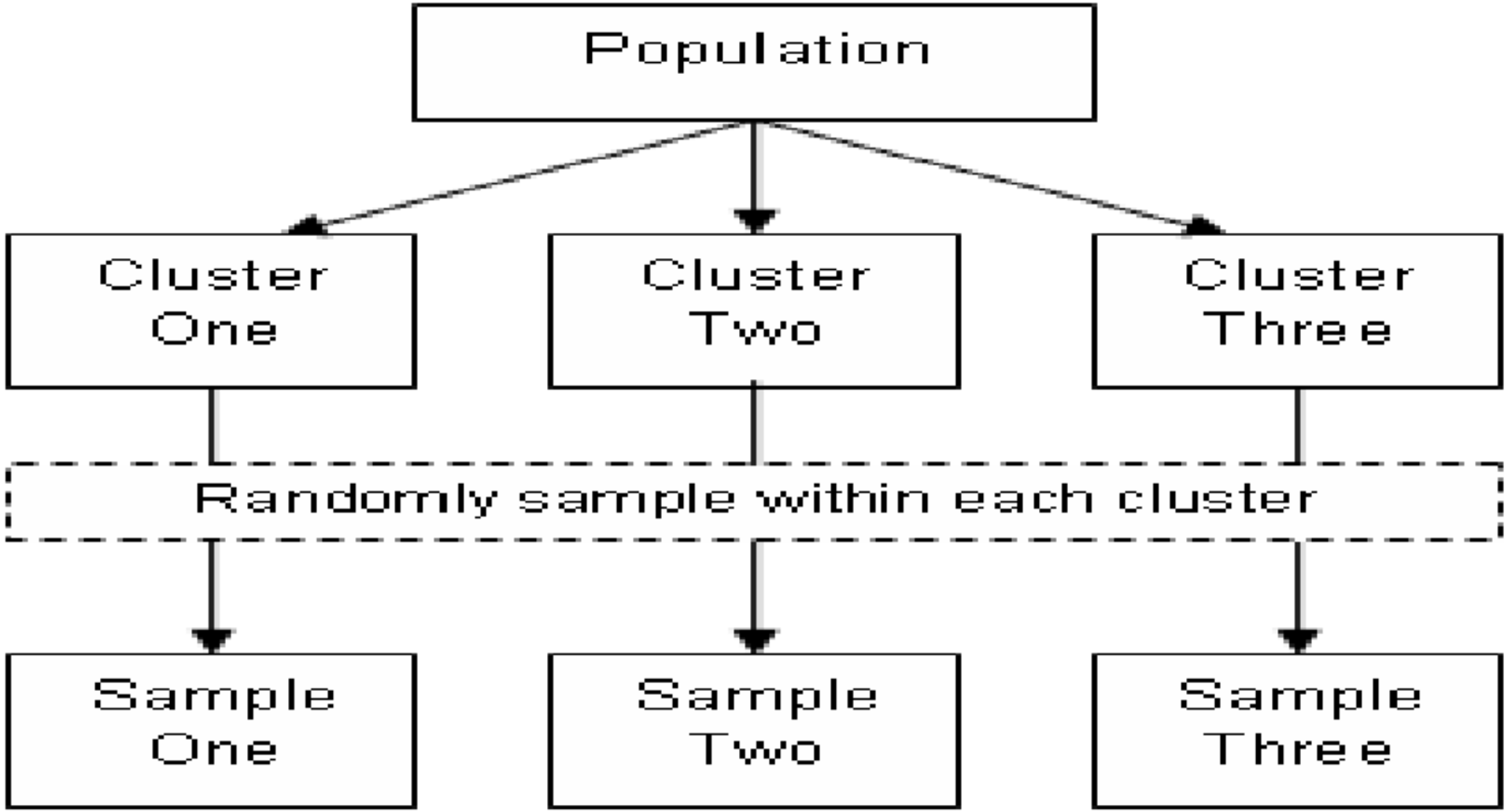
- Very useful when population is spread over large geographic area.
- Convenient and expedient
- Does not need names of everyone in population.
- Reduced cost

Disadvantages:

- Representation is likely to become an issue
- Less precise than random sampling.

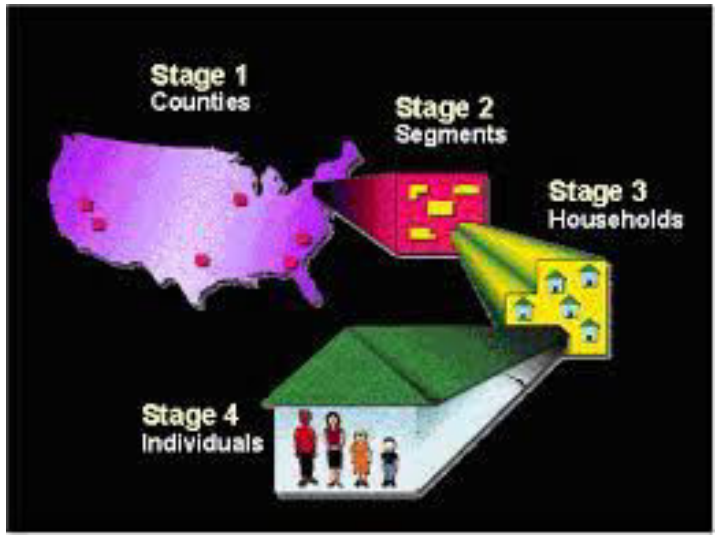
If clusters (or groups/subdivisions) are geographic areas, then this is called Area Sampling.

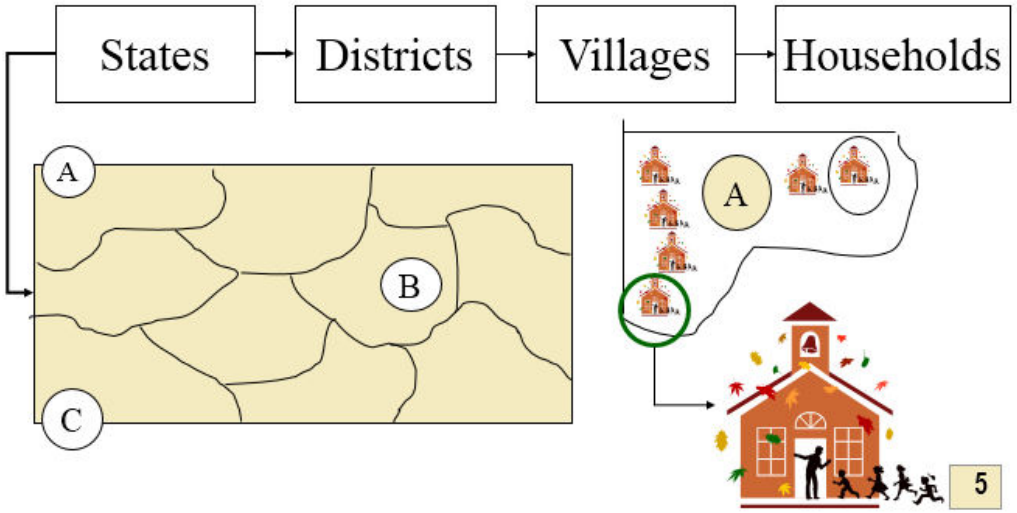
CLUSTER SAMPLING



MULTISTAGE SAMPLING

- Selects samples using more than 1 sampling technique.
- Complex - Hence rarely used.
- Requires lot of effort, time & cost.



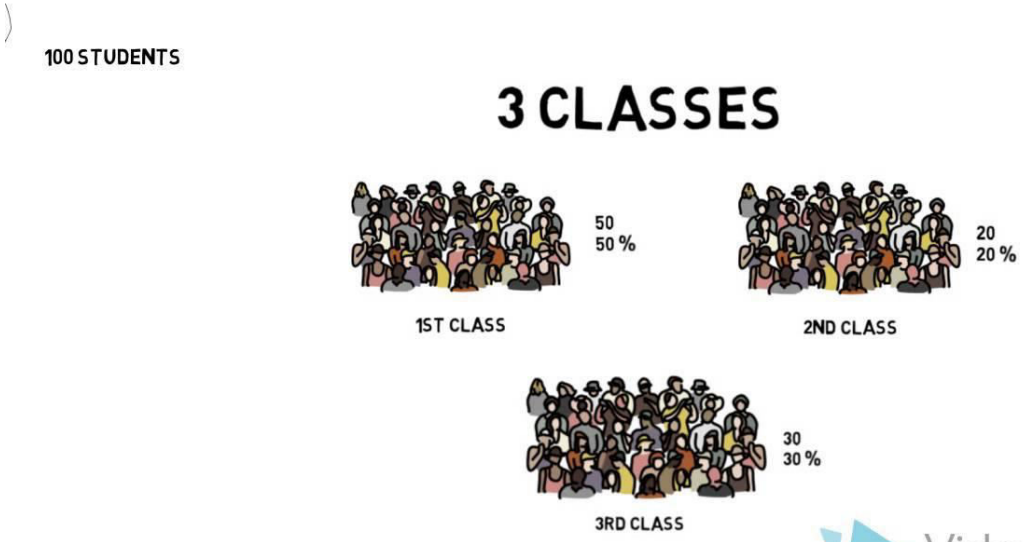


SEQUENTIAL SAMPLING

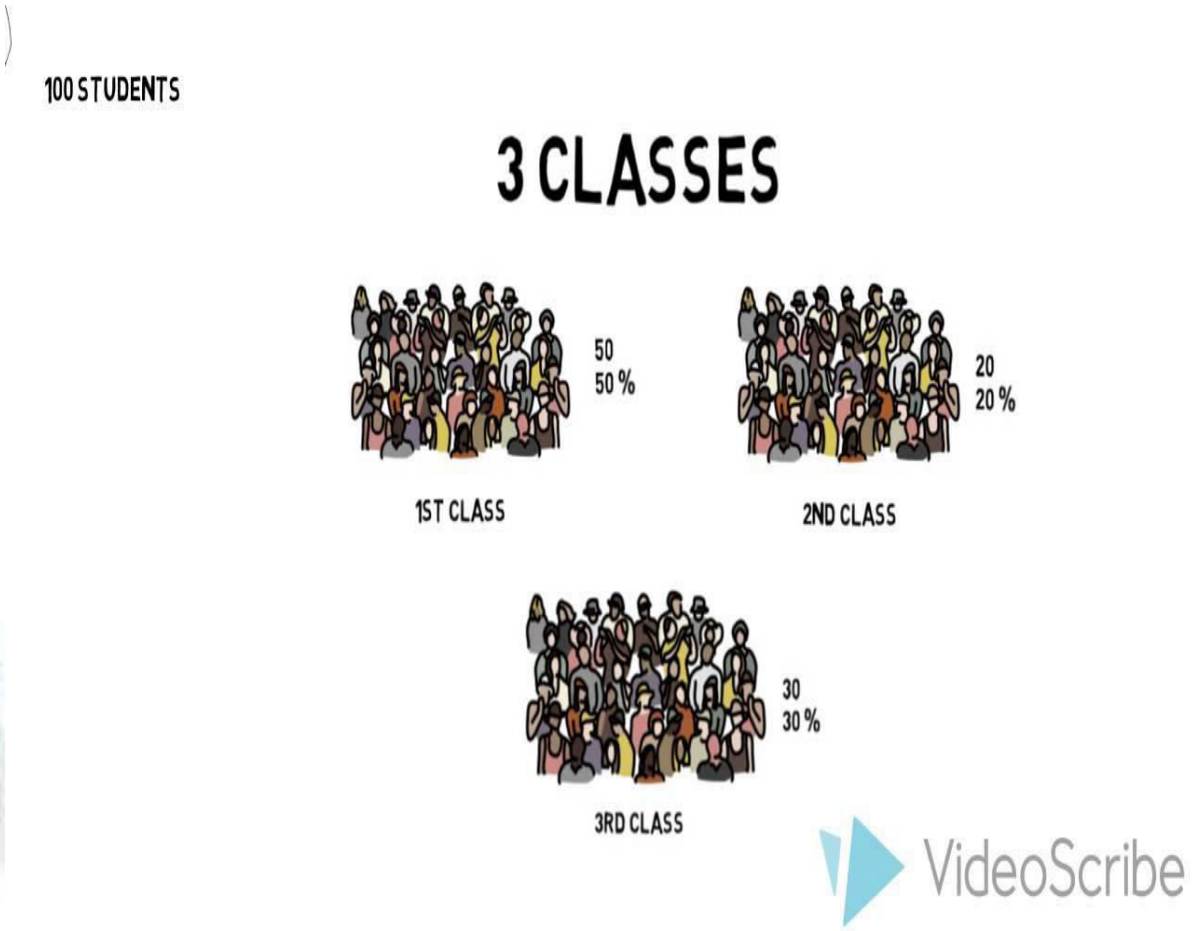
- Complex Sample design
- Size is not fixed in advance.
- Adopted in case of acceptance sampling.
    - When a particular lot to be accepted / rejected on basis of single sample - single sampling
    - When decision taken on basis of 2 samples - double sampling
- When number of samples are more than two, but neither certain nor decided in advance - Sequential Sampling

QUOTA SAMPLING

- This is a type of non probability sampling.
- Population is divided into mutually exclusive sub groups as in stratified sampling.
- Judgement is used to select the subject or unit from each segment based on specified portion.
- Quota sampling is judgement samples rather than random samples.



QUOTA SAMPLING



THANK YOU

Department of Computer Science and Engineering

RESEARCH METHODOLOGY

Methods of Data Collection

Dr.Roopa Ravish
Department of Computer Science and Engineering

RESEARCH METHODOLOGY

Ch. 6: Methods of Data Collection

Department of Computer Science and Engineering

CONTENTS

1) Introduction
2) Collection of Primary Data
3) Types of Collection of Primary Data
    a)Observation Method
    b)Interview Method
    c)Through Questionnaires
    d)Through Schedules
        ◆Differences
a) Collection of Secondary Data
b) Selection Criteria

DATA COLLECTION

Primary data Collection :

- Fresh
- First Time
- Original in Character

Secondary data Collection :

- Already collected by some one
- Already passed through statistical process

SURVEY Vs EXPERIMENT

We collect primary data during the course of doing experiments in an experimental research but in case we do research of the descriptive type and perform surveys, whether sample surveys or census surveys, then we can obtain primary data either through observation or through direct communication with respondents in one form or another or through personal interview

SURVEY Vs EXPERIMENT

Survey	Experiment
Descriptive research	Experimental research
Large number of data	Small Number of data
No manipulation	Deliberate Manipulation
Social and Behavioral Studies	Physical and Natural Studies
Data collection by observation, interview, group discussion	Data collection by repeated reading
Relationship between data and unknown can be studied through survey	Determine relationship between them
Casual analysis	Coorelation analysis

COLLECTION OF PRIMARY DATA

1. Observation Method
2. Interview Method
    a. Personal Interview
    b. Telephonic Interview
3. Questionnaires
4. Schedules
5. Other methods

OBSERVATION Method

- Related to Behavioral Sciences
- Non scientific method – To observe things around us
- Scientific tool -
  - systematically planned and recorded,
  - subjected to checks and
  - Controls on validity and reliability.
- Information is sought by way of investigators own direct observation w/o asking respondents

Advantages :

1. Subjective bias is eliminated ; if observation done accurately.
2. Relates to current happenings
3. Independent of respondents

Limitations:

1. Expensive method
2. Limited information is provided.
3. Unforeseen factors interfere
4. People may not be accessible

OBSERVATION

While using this method researcher should keep following in mind

- What should be observed?
- How should observation be recorded?
- How can accuracy be ensured?

Few Terminologies

- Structured Observation
  - Units, style, standardised conditions, pertinent data selection
  - Descriptive studies
- Unstructured Observation

Exploratory studies when observation
is to take place without these characteristics to be thought of in advance,

- Participant Observation
- Non Participant Observation (Disguised Observation)
- Controlled and Non Controlled Observation.

OBSERVATION

- Participant Observation and Non Participant Observation (Disguised Observation)

If the observer observes by making himself, more or less, a member of the group he is observing so that he can experience what the members of the group experience, the observation is called as the *participant observation.* But when the observer observes as a detached emissary without any attempt on his part to experience through participation what others feel, the observation of this type is often termed as *non-participant observation.*

OBSERVATION

There are several merits of the participant type of observation:

(i) The researcher is enabled to record the natural behaviour of the group.

(ii) The researcher can even gather information which could not easily be obtained if he observes in a disinterested fashion.

(iii) The researcher can even verify the truth of statements made by informants in the context of a questionnaire or a schedule.

Demerits of this type of observation :

The observer may lose the objectivity to the extent he participates emotionally;

Problem of observation-control is not solved;

It may narrow-down the researcher's range of experience.

OBSERVATION

- Controlled and Non Controlled Observation.

If the observation takes place in the natural setting, it may be termed as uncontrolled observation.
When observation takes place according to definite pre-arranged plans, involving experimental procedure, the same is then termed controlled observation.
In non-controlled observation, no attempt is made to use precision instruments.
The major aim of this type of observation is to get a spontaneous picture of life and persons. It has a tendency to supply naturalness and completeness of behaviour, allowing sufficient time for observing it.
But in controlled observation, we use mechanical (or precision) instruments as aids to accuracy and standardisation. Such observation has a tendency to supply formalised data upon which generalisations can be built with some degree of assurance.

OBSERVATION

- Controlled and Non Controlled Observation.

The main pitfall of non-controlled observation is that of subjective interpretation. There is also the danger of having the feeling that we know more about the observed phenomena than we actually do.

Generally, controlled observation takes place in various experiments that are carried out in a laboratory or under controlled conditions, whereas uncontrolled observation is resorted to in case of exploratory researches.

INTERVIEW METHOD

Presentation of Oral Verbal Stimuli and reply in oral verbal responses.

2 types

- Personal Interview

- Telephonic Interview

PERSONAL INTERVIEW METHOD

- Face to Face contact with person / persons.
- Direct personal investigation or indirect oral investigation
- DPI- Interviewer collect information personally from sources concerned.
- He has to be on spot and has to meet people from whom data have to be collected.

Structured Interview:

- Method of collecting information through personal interview usually done is structured way.
- Predetermined set of questions.
- Rigid procedure asking the questions in form and order prescribed.

Unstructured Interview

- Flexibility approach of questioning.
- Do not follow predetermined questions and standards
- Interviewer allowed much greater freedom to ask, ask supplementary questions, omit certain questions.
- He may change sequence of questions

Focused Interview
Clinical Interview
Non Directive Interview

PERSONAL INTERVIEW METHOD Merits

(i) More information and that too in greater depth can be obtained.
(ii) Interviewer by his own skill can overcome the resistance, if any, of the respondents; the interview method can be made to yield an almost perfect sample of the general population.
(iii) There is greater flexibility under this method as the opportunity to restructure questions is always there, specially in case of unstructured interviews.
(iv) Observation method can as well be applied to recording verbal answers to various questions.
(v) Personal information can as well be obtained easily under this method.
(vi) Samples can be controlled more effectively as there arises no difficulty of the missing returns; non-response generally remains very low.
(vii) The interviewer can usually control which person(s) will answer the questions. This is not possible in mailed questionnaire approach. If so desired, group discussions may also be held.

PERSONAL INTERVIEW METHOD Merits

(viii) The interviewer may catch the informant off-guard and thus may secure the most spontaneous reactions than would be the case if mailed questionnaire is used.

(ix) The language of the interview can be adopted to the ability or educational level of the person interviewed and as such misinterpretations concerning questions can be avoided.

(x) The interviewer can collect supplementary information about the respondent's personal characteristics and environment which is often of great value in interpreting results.

DEMERITS OF INTERVIEW METHOD

(i) It is a very expensive method, specially when large and widely spread geographical sample is taken.

(ii) There remains the possibility of the bias of interviewer as well as that of the respondent; there also remains the headache of supervision and control of interviewers.

(iii) Certain types of respondents such as important officials or executives or people in high income groups may not be easily approachable under this method and to that extent the data may prove inadequate.

(iv) This method is relatively more-time-consuming, specially when the sample is large and recalls upon the respondents are necessary.

DEMERITS OF INTERVIEW METHOD

(v) The presence of the interviewer on the spot may over-stimulate the respondent, sometimes even to the extent that he may give imaginary information just to make the interview interesting.

(vi) Under the interview method the organisation required for selecting, training and supervising the field-staff is more complex with formidable problems.

(vii) Interviewing at times may also introduce systematic errors.

(viii) Effective interview presupposes proper rapport with respondents that would facilitate free and frank responses. This is often a very difficult requirement.

PREREQUISITES & BASIC TENETS OF INTERVIEW METHOD

- Interviewer should be carefully selected, trained and briefed.
- They should be honest, sincere and hardworking, must possess technical competence and necessary practical experience.
- Field checks should be made – to ensure interviewer is neither cheating nor deviating from instructions given.

TELEPHONIC INTERVIEW

- Method of collecting information contacting respondents over telephone.
- Not widely used, but plays important role in industry surveys.

Merits:

1. It is more flexible in comparison to mailing method.
2. It is faster than other methods i.e., a quick way of obtaining information.
3. It is cheaper than personal interviewing method; here the cost per response is relatively low.
4. Recall is easy; callbacks are simple and economical.
5. There is a higher rate of response than what we have in mailing method; the non-response is generally very low.
6. Replies can be recorded without causing embarrassment to respondents.

TELEPHONIC INTERVIEW

Merits:

7. Interviewer can explain requirements more easily.
8. At times, access can be gained to respondents who otherwise cannot be contacted for one reason or the other.
9. No field staff is required.
10. Representative and wider distribution of sample is possible.

TELEPHONIC INTERVIEW

Demerits:

1. Little time is given to respondents for considered answers; interview period is not likely to exceed five minutes in most cases.
2. Surveys are restricted to respondents who have telephone facilities.
3. Extensive geographical coverage may get restricted by cost considerations.
4. It is not suitable for intensive surveys where comprehensive answers are required to various questions.
5. Possibility of the bias of the interviewer is relatively more.
6. Questions have to be short and to the point; probes are difficult to handle.

COLLECTION OF DATA THROUGH QUESTIONNAIRES

- Big enquiries – economic and business surveys.
- Being adopted – private individuals, research workers, private and public organizations, governments.
- Questions are sent (by post) to persons concerned and required to answer and return.
- Consists of number of questions printed or typed in definite order on a form/s.
- Questionnaire is mailed to respondent/s and who is expected to read and understand the questions and write down their reply in space meant for the purpose.

QUESTIONNAIRE

Merits

1. Low cost even over large geographic area.
2. Free from bias of interviewer
3. Respondents have adequate time to respond
4. Respondents who are not easily approachable can also be reached conveniently.
5. Large samples can be made use of, results can be reliable

Demerits

1. Low rate of return
2. Used when respondents are educated and cooperating.
3. Control over questionnaire may be lost when sent.
4. Inbuilt inflexibility – amending approach once dispatched.
5. Difficult to know whether willing respondents are truly representative.
6. Slowest of all methods.
7. Incomplete/ambiguous answers – hard to interpret.

QUESTIONNAIRE

Pilot Study may be required to test the questionnaire – rehearsal of main survey:

- Weakness of questionnaire may be identified
- Omission of questions
- Modification (reordering/rewording/rephrasing) of questions may be required – wrong interpretation of question by respondent

Main aspects of questionnaire:

a)General form

b)Question sequence

c)Question formulation and wordings

- Very clear in construction – simple, easily understood
- Open ended questions
- Multiple choice questions, True/False

Essentials of a good questionnaire:

Main aspects of a questionnaire:

1. *General form:*
2. *Question sequence:*
3. *Question formulation and wording:*

Essentials of a good questionnaire:

To be successful, questionnaire should be short and simple i.e., the size of the questionnaire should be kept to the minimum.

Questions should proceed in logical sequence moving from easy to more difficult questions.

Personal and intimate questions should be left to the end.

Technical terms and vague expressions capable of different interpretations should be avoided in a questionnaire.

Questions may be dichotomous (yes or no answers), multiple choice (alternative answers listed) or open-ended. The latter type of questions are often difficult to analyse and hence should be avoided in a questionnaire to the extent possible.

Essentials of a good questionnaire:

There should be some control questions in the questionnaire which indicate the reliability of the respondent.
For instance, a question designed to determine the consumption of particular material may be asked first in terms of financial expenditure and later in terms of weight. The control questions, thus, introduce a cross-check to see whether the information collected is correct or not.
Questions affecting the sentiments of respondents should be avoided.

Essentials of a good questionnaire:

Adequate space for answers should be provided in the questionnaire to help editing and tabulation.

There should always be provision for indications of uncertainty, e.g., “do not know,” “no preference” and so on. Brief directions with regard to filling up the questionnaire should invariably be given in the questionnaire itself.

Finally, the physical appearance of the questionnaire affects the cooperation the researcher receives from the recipients and as such an attractive looking questionnaire, particularly in mail surveys, is a plus point for enlisting cooperation.

The quality of the paper, along with its colour, must be good so that it may attract the attention of recipients.

SCHEDULES

- The schedule is another data collection technique containing statements, questions, and blank spaces to fill up the answers given by the respondents to the enumerator or interviewer.

- A schedule is a structure of a set of questions on a given topic which are asked by the interviewer or investigator personally. ... It contains direct questions as well as questions in tabular form. Schedule include open-ended questions and close-ended questions.

- Schedules are filled by enumerators who are appointed for this purpose.
- These enumerators go to respondents, ask listed questions and record the replies.
- In few instances schedules may be handed over to respondents and enumerators may help them in recording answers.
- Enumerators should explain objectives of investigation and remove difficulty giving appropriate clarifications.

SCHEDULES

- **Benefits**:
- It is a useful method in case the informants are illiterate.
- The researcher can overcome the problem of non-response as the enumerators go personally to obtain the information.
- It is very useful in extensive studies and
- Can obtain more reliable data

- **Demerits**
  - Very expensive
    - Enumerator selection
    - Enumerators to be trained extensively

Differences between Questionnaires and Schedules:

Both questionnaires and schedules are popularly used methods of collecting data in research surveys. There is much resemblance in the nature of these two methods. From a practical point of view, the two methods can be taken to be the same. But from a technical point of view there is difference between the two.

The important points of difference are as under:

Differences between Questionnaires and Schedules:

#	Questionnaires	Schedules
1	• Filled out by the respondent.	Generally filled out by the research worker or the enumerator.
2	Relatively cheap and economical; no field staff required.	Relatively more expensive; appointing enumerators and in imparting training to them; preparing schedules also has costs.
3	Non-response is usually high; Bias due to non-response often remains indeterminate.	Generally Non-response very low; Since filled by enumerators who get answers to all questions; Danger of interviewer bias and cheating exists.
4	Not always clear as to who replies, respondent or someone on his/her behalf.	Identity of respondent is known to enumerator.
5	Personal contact is generally not possible.	Direct personal contact is established with respondents by enumerator.

Differences between Questionnaires and Schedules:

#	Questionnaires	Schedules
6	Personal contact is generally not possible.	Direct personal contact is established with respondents.
7	Can be used only when respondents are literate and cooperative.	Respondents may be illiterate.
8	Wider and more representative distribution of sample	Difficulty in sending enumerators over a relatively wider area.
9	Risk of collecting incomplete and wrong information is relatively more particularly when people are unable to understand questions properly.	Relatively more accurate; Generally complete and accurate as enumerators can remove the difficulties, if any.
10	Success lies more on the quality of the questionnaire itself.	Depends upon the honesty and competence of enumerators.
11	Physical appearance of questionnaire must be quite attractive.	Not Applicable.
12	Not possible.	Observation method can also be used.

SECONDARY DATA COLLECTION

- Data already available – already collected and analysed by someone else.
- Researcher should look into sources from where s/he can obtain.
- 2 forms of secondary data:
  - Published
  - Unpublished.

SECONDARY DATA COLLECTION

Published -

a) Publication in form of central, state and local government.
b) Publication of foreign govt and international bodies.
c) Technical and trade journals
d) Books, magazines and newspapers
e) Reports and publications of various business associations, industries, banks, stock exchanges, etc.
f) Reports prepared by universities, scholars, economists
g) public records and statistics, historical documents,
h) websites - Ex : RBI, NSE, etc.

Unpublished

- diaries, letters, biographies, autobiographies, also made available with scholars and workers, trade associations, etc.

SECONDARY DATA COLLECTION

Caution before using secondary data to see following characteristics:

1. Reliability of data:
    - Who collected the data?
    - What were the sources of data?
    - Were they collected using proper methods?
    - At what time were they collected?
    - Was it achieved?

2. Suitability of data:
    - Data suitable for one enquiry may not be suitable for another. Hence if the data is found unsuitable it should not be used.

3. Adequacy of data:
    - If the level of accuracy is found to be inadequate, it should not be considered for research.

THANK YOU

Priya Badarinath , Raghu B. A, Roopa Ravish
Department of Computer Science and Engineering

priyab@pes.edu, raghubarao@pes.edu,
rooparavish@pes.edu