Unit 2
Research Design

Research Design

Research Design

- Task of defining the research problem is the preparation of the research project, popularly known as the “research design".
- Decisions regarding what, where, when, how much, by what means concerning an inquiry or a research study constitute a research design.
- A plan or strategy for conducting the research.
- A research design is one that minimizes bias and maximizes the reliability of the data.
- It also yields maximum information, gives minimum experimental error, and provides different aspects of a single problem.
- A research design depends on the purpose and nature of the research problem. Thus, one single design cannot be used to solve all types of research problem, i.e., a particular design is suitable for a particular problem.

Need of research design

- Smooth sailing of research operations.
- Stands for advance planning of the methods.
- Great bearing on reliability of the results.
- Helps to give directions.
- Helps in decision making.
- Research design prevent blind searching.
- Helps researchers to anticipate potential problems in collecting data etc.

Features of a Good Design

- The means of obtaining information
- The availability and skills of the researcher and his staff, if any;
- The availability of time and money for the research work.
- It should be flexible enough to consider different aspects of the study in case of exploratory.
- The design should be accurate with minimum bias in case of accurate description
- Control of extraneous variables
- Statistical correctness for testing hypothesis

Concepts relating to research design

- Dependent and independent variables
- Extraneous variable
- Control
- Experimental and non-experimental hypothesis- testing research
- Experimental and control groups
- Treatments

Variables

- Any characteristic which is subject to change and can have more than one value such as age, intelligence, motivation, gender, etc.
- Types of variables
- **Independent vs. Dependent vs. Controlled Variables**
- **Categorical vs. Continuous Variables**
- **Quantitative vs. Qualitative Variables**

Dependent Variable

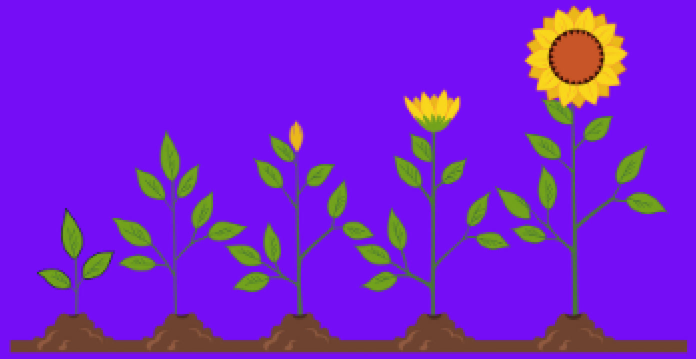
[di-PEN-duhnt VAIR-ee-uh-bl]

Definition:

A dependent variable is the measured outcome in an experiment, influenced by changes in the independent variable.

Examples:

Plant Growth Experiment



Dependent Variable: Height of the plants.

Student Performance Study

Dependent Variable: Student test scores.

Examples.com

Dependent Variable

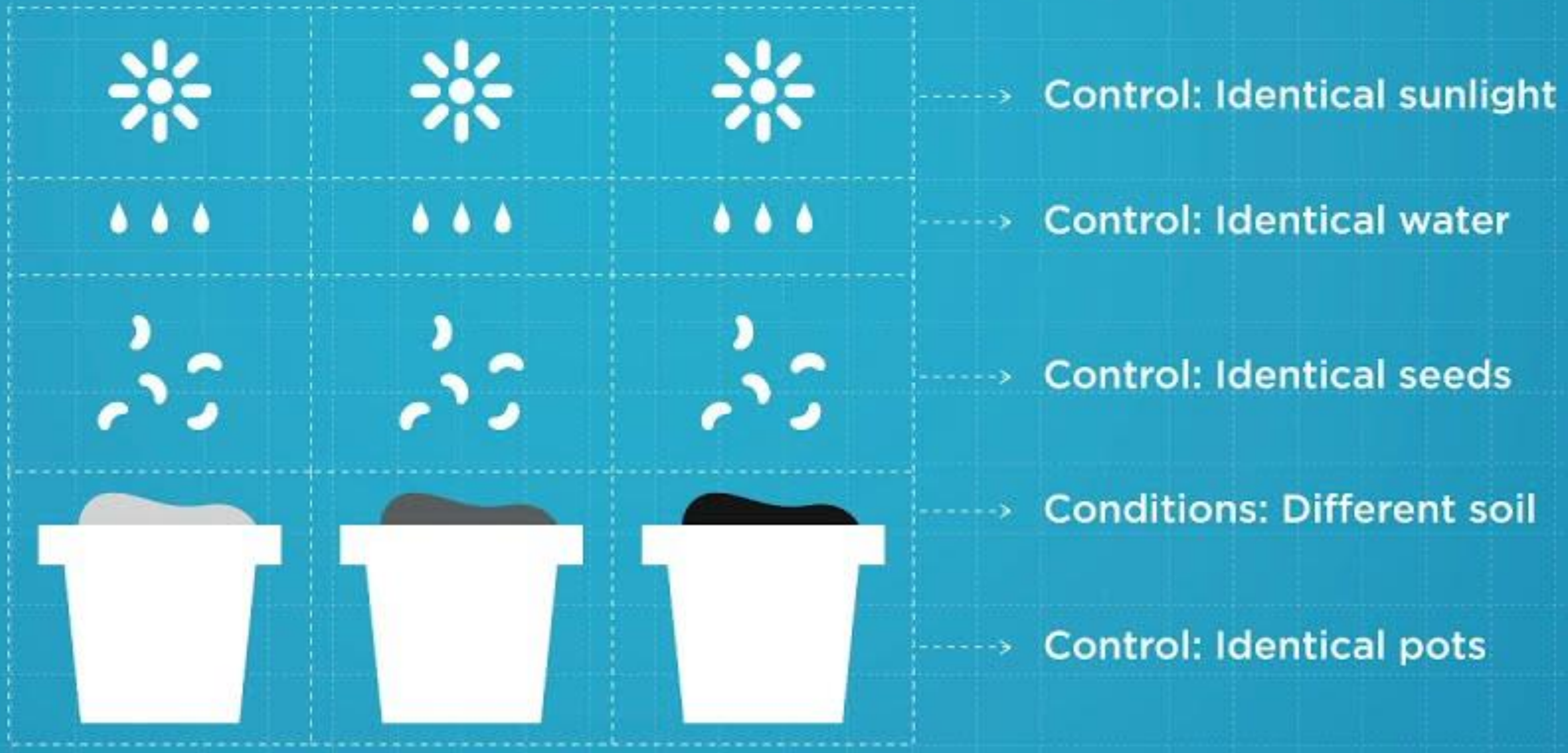
- Variable affected by the independent variable
- It responds to the independent variable.
- In an experiment that which is supposed to be changed by the independent.

Independent Variable

- Variable that is presumed to influence other variable
- It is the presumed cause, whereas the dependent variable is the presumed effect.
- In an experiment that which is supposed to be manipulated by you.
- The variable manipulated by the experimenter.

Experimental Research: Extraneous Variables

Extraneous Variables are variables other than the ones you are interested that may be interfering with results.



Extraneous variable

- Independent variable that are not related to the purpose of the study, but may affect the dependent variable are termed as extraneous variables.
- Whatever effect is notices on dependent variable as a result of extraneous variable is technically described as an 'experimental error'.

Control

- One of the important characteristics of a good research design is to minimize the influence or effect of extraneous variable.

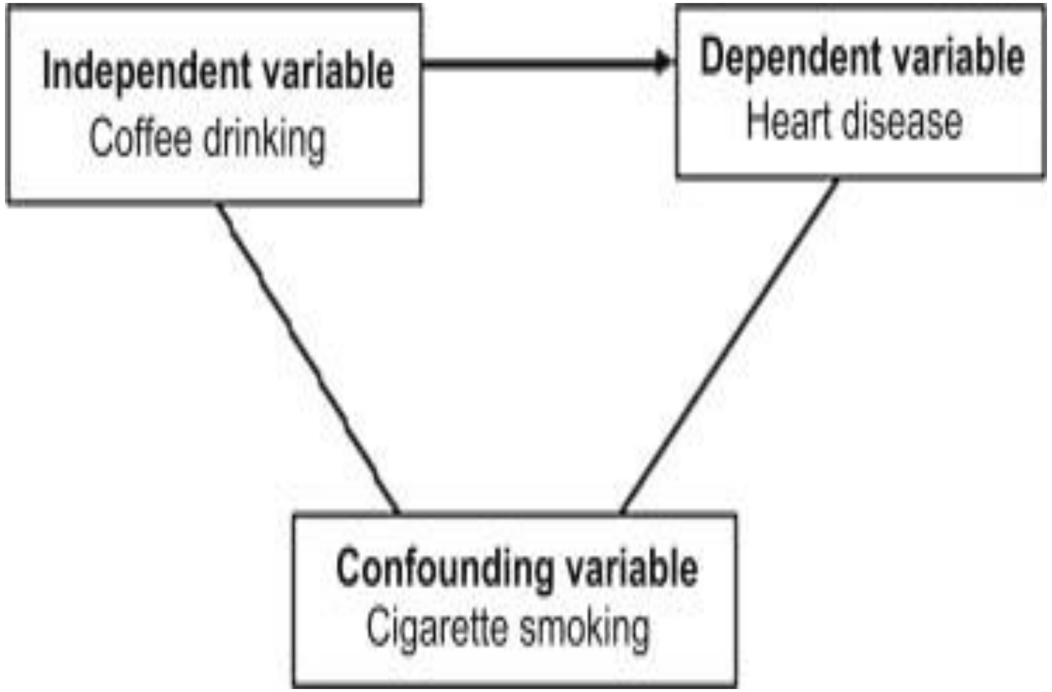
Experimental and non experimental hypothesis testing

- Research in which the independent variable is manipulated is termed experimental hypothesis –testing research.
- Research in which an independent variable is not manipulated is called non-experimental hypothesis.

Confounded Relationship

Confounding relationship means the distortion of the association between the independent and dependent variables because a third variable is independently associated with both.

A causal relationship between two variables is often described as the way in which the independent variable affects the dependent variable.



Key Characteristics of Confounding Variables

1. Correlates With the Independent Variable
2. Associates With the Dependent Variable
3. Introduces Noise into the Relationship
4. Lacks Relation to the Research Question

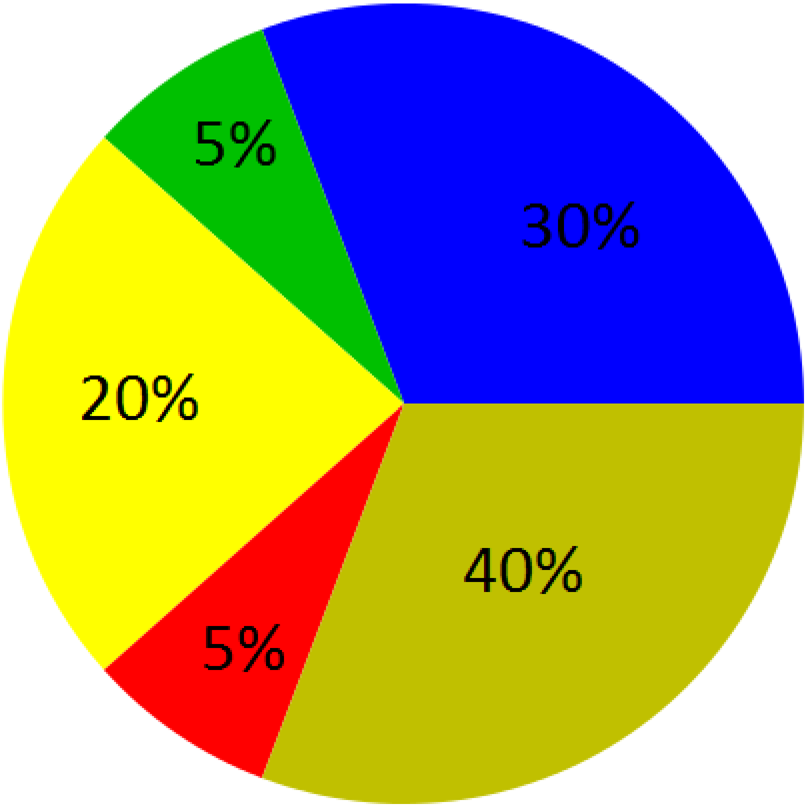
Experimental and control group

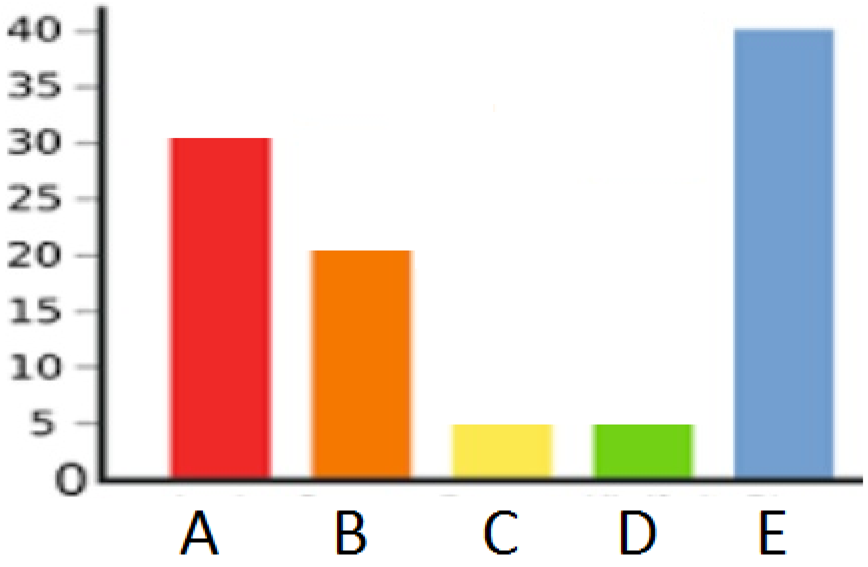
- In an experimental hypothesis –testing research when a group is exposed to usual conditions, it is termed as control group.
- When a group is exposed to some novel or special condition it is termed as experimental group.

Treatments

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

Categorical Data





Categorical vs. Continuous Variables

- *Categorical variables are variables that can take on* specific values only within a defined range of values like gender, marital status
- consisting of discrete, mutually exclusive categories, such as “male/female,” “White/Black,” etc
- *Continuous variables are variables that can theoretically take on any* value along a continuum like age, income weight, height etc..
- When compared with categorical variables, continuous variables can be measured with a greater degree of precision.
- The choice of which statistical tests will be used to analyze the data is partially dependent on whether the researcher uses categorical or continuous variables.
- Certain statistical tests are appropriate for categorical variables, while other statistical tests are appropriate for continuous variables.
- As with many decisions in the research-planning process, the choice of which type of variable to use is partially dependent on the question that the researcher is attempting to answer.

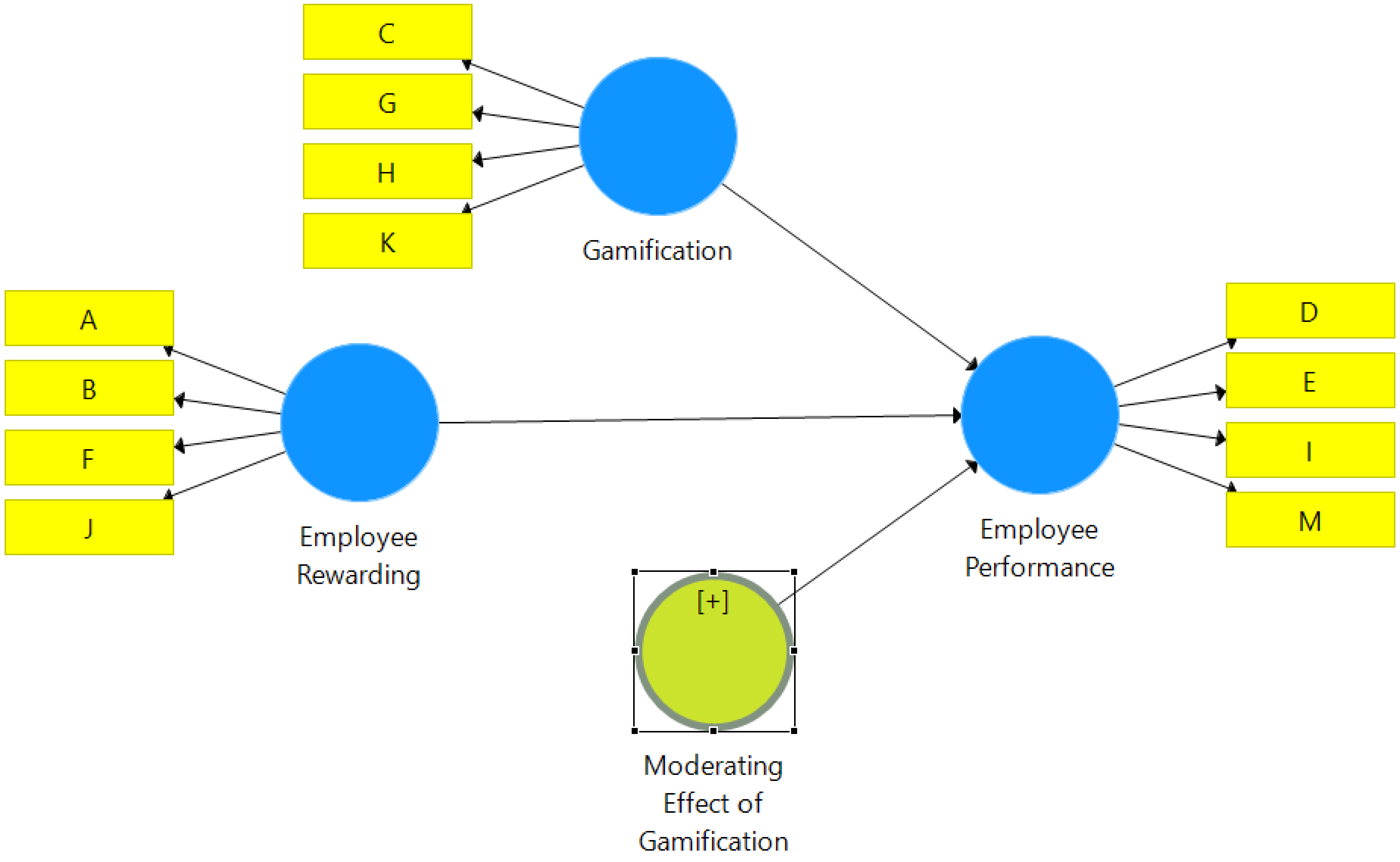
Quantitative vs. Qualitative Variables

- *Qualitative variables are variables that vary in kind, like* "attractive" or "not attractive," "helpful" or "not helpful," or "consistent" or "not consistent"

- *Quantitative variables are those that vary in amount like height, weight, salary etc*

Hypothesis

- The research hypothesis is a predictive statement that relates an independent variable to dependent variable.
- A hypothesis may be defined as a proposition or set of proposition set forth as an explanation for occurrence of some specified group of phenomena either asserted merely as a provisional conjucture to guide some investigation or accepted as highly probable in the light of established facts


C
G
H
K
Gamification
A
B
F
J
Employee
Rewarding
D
E
I
M
Employee
Performance
[+]
Moderating
Effect of
Gamification

Purpose

- Guides/gives direction to the study/investigation
- Defines Facts that are relevant and not relevant
- Suggests which form of research design is likely to be the most appropriate
- Provides a framework for organizing the conclusions of the findings
- Limits the research to specific area
- Offers explanations for the relationships between those variables that can be empirically tested
- Furnishes proof that the researcher has sufficient background knowledge to enable her/him to make suggestions in order to extend existing knowledge
- Structures the next phase in the investigation and therefore furnishes continuity to the examination of the problem

Characteristics of hypothesis

(i) Hypothesis should be clear and precise. the inferences drawn on its basis cannot be taken as reliable.

(ii) Hypothesis should be capable of being tested. A hypothesis “is testable if other deductions can be made from it which, in turn, can be confirmed or disproved by observation.”

(iii) Hypothesis should state the relationship between variables if it happens to be a relational hypothesis.

(iv) Hypothesis should be limited in scope and must be specific. A researcher must remember that narrower hypotheses are generally more testable and he should develop such hypotheses

Characteristics of hypothesis

(v) Hypothesis should be stated as far as possible in simple terms so that the same is easily understandable..

(vi) Hypothesis should be consistent with most known facts i.e., it must be consistent with a substantial body of established facts.

(vii) Hypothesis should be amenable to testing within a reasonable time. One should not use even an excellent hypothesis if the same cannot be tested in a reasonable time for one cannot spend a lifetime collecting data to test it.

(viii) Hypothesis must explain the facts that gave rise to the need for explanation. This means that by using the hypothesis plus other known and accepted generalizations, one should be able to deduce the original problem condition.

Experimental and non-experimental hypothesis testing

- When a group is exposed to usual conditions, it is termed as a control group.
- But when the group is exposed to be some special condition, it is termed as Experimental group

Categorizing Hypotheses

Can be categorized in different ways

1. Based on their *formulation*

- Null Hypotheses and Alternate Hypotheses

2. Based on direction

- Directional and Non-directional Hypothesis

1. Null Hypotheses and Alternate Hypotheses

- Null hypothesis always predicts that
  - no differences between the groups being studied (e.g., experimental vs. control group) or
  - no relationship between the variables being studied

- By contrast, the alternate hypothesis always predicts that there will be a difference between the groups being studied (or a relationship between the variables being studied)

H0: There is no difference in the salary of factory workers based on gender.

Alternative Hypothesis:

Ha: Male factory workers have a higher salary than female factory workers.

2. Directional Hypothesis and Non-directional Hypothesis

- Simply based on the wording of the hypotheses we can tell the difference between directional and non-directional
  - If the hypothesis simply predicts that there will be a difference between the two groups, then it is a ***non-directional hypothesis***. It is non-directional because it predicts that there will be a difference but does not specify how the groups will differ.
  - If, however, the hypothesis uses so-called comparison terms, such as "greater," "less," "better," or "worse," then it is a ***directional hypothesis***. It is directional because it predicts that there will be a difference between the two groups and it specifies how the two groups will differ

Example: Therapy decreases depression.

Example : The memory pill will lead to some type of difference in memory performance

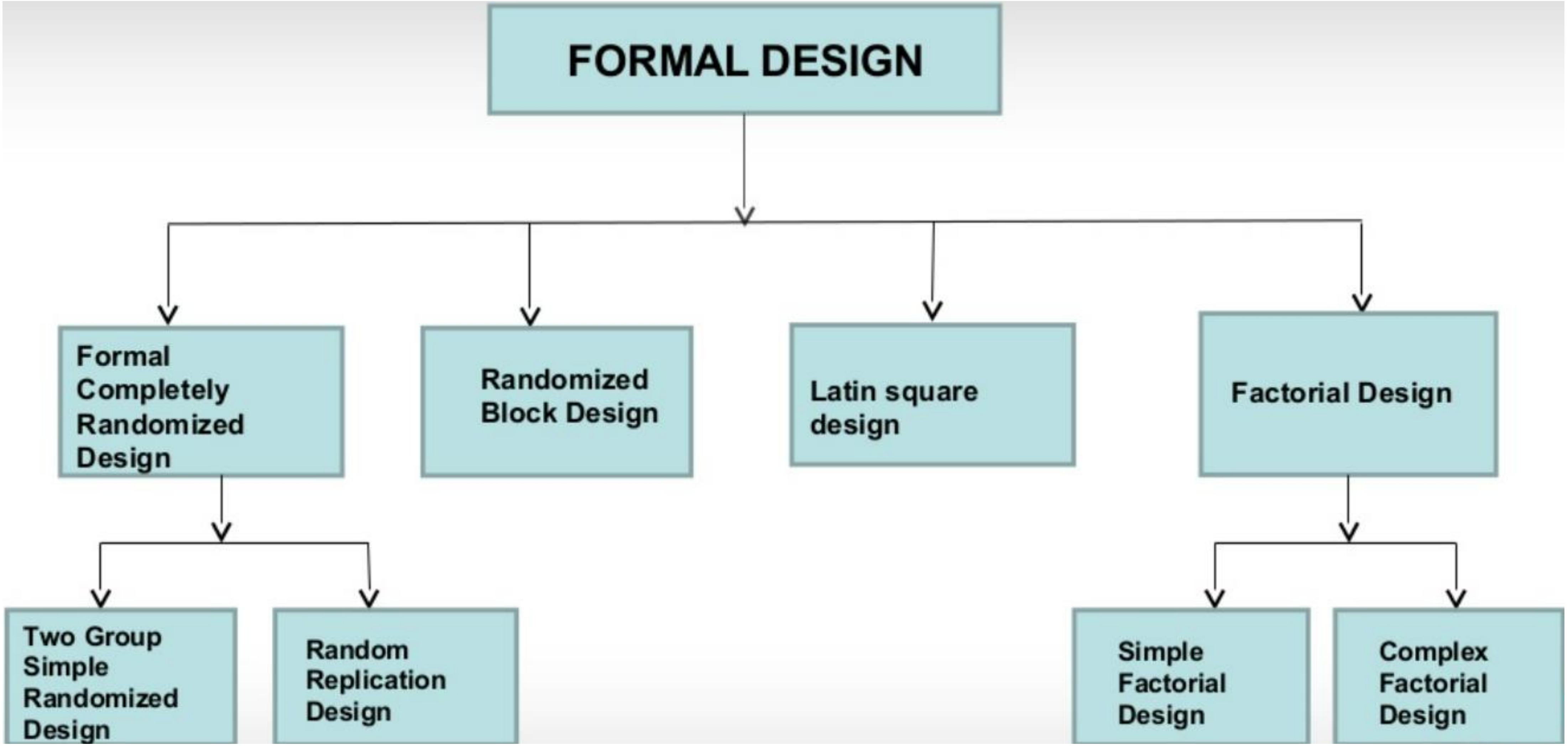
Example : The memory pill will increase or improve memory performance

Experimental Design

- Experimental design methods allow the experimenter to understand better and evaluate the factors that influence a particular system by means of statistical approaches.
- Such approaches combine ***theoretical knowledge of experimental designs*** and a ***working knowledge*** of the particular factors to be studied.

Principles of Experimental Design

- **Control** – make conditions as similar as possible for all treatment groups (aside from the actual treatments).
  - *If we observe a difference between groups, we want to know that it is a result of the treatment(s)!*
- **Randomization** – the use of chance to assign subjects/units to treatments
  - *This helps create roughly equivalent groups of experimental units by balancing the effects of lurking variables that aren't controlled on the treatment groups.*
- **Replication** of the experiment on many subjects/in different locations/etc.



Single factor/Single independent variable

Multi factors/Multiple independent variables

Factorial Design

- 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:

1. simple factorial design and
2. complex factorial designs.

Simple Factorial Design

- we consider the effects of varying two factors on the dependent variable.
- Simple factorial design is also termed as a ‘two-factor-factorial design’, whereas complex factorial design is known as ‘multifactor-factorial design.’
- Simple factorial design may either be a 2 × 2 simple factorial design, or it may be, say, 3 × 4 or 5 × 3 or the like type of simple factorial design.

Illustration 1: (2 × 2 simple factorial design).

A 2 × 2 simple factorial design can graphically be depicted as follows:

2 × 2 SIMPLE FACTORIAL DESIGN

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

Simple Factorial Design

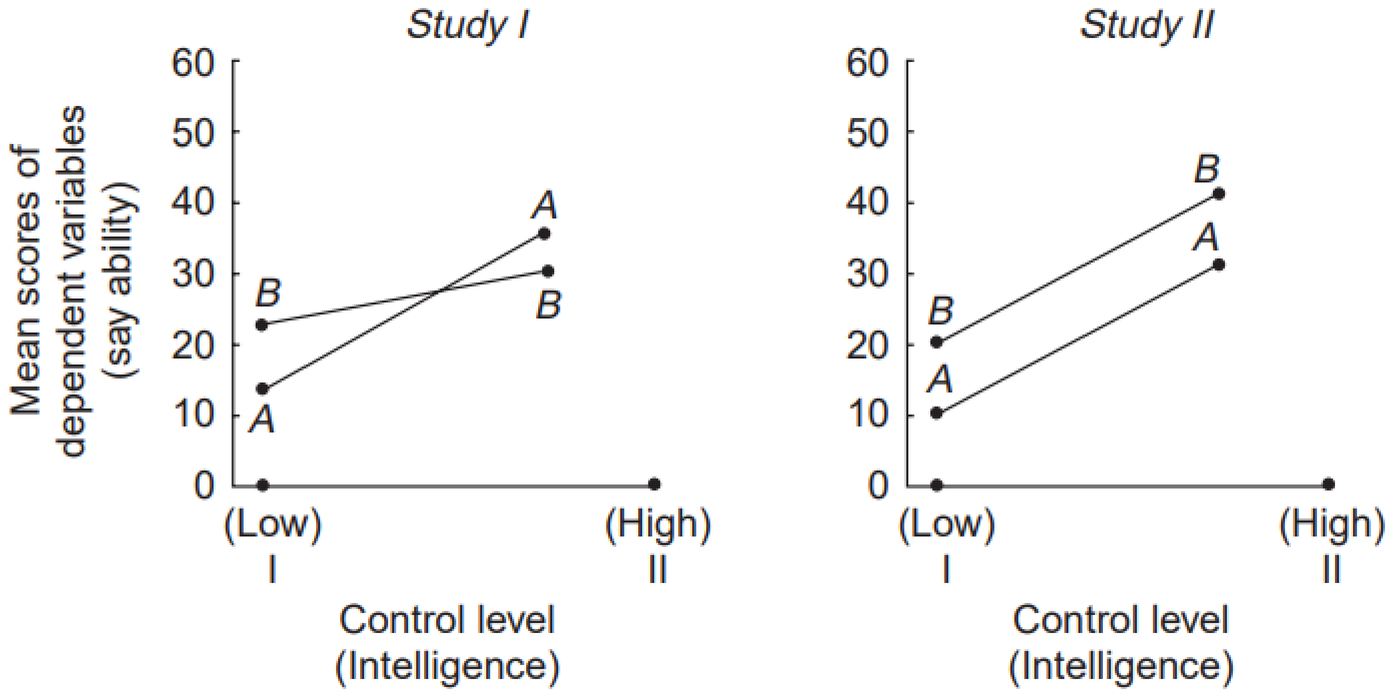
STUDY I DATA

		Training		
		Treatment A	Treatment B	Row Mean
Control (Intelligence)	Level I (Low)	15.5	23.3	19.4
	Level II (High)	35.8	30.2	33.0
	Column mean	25.6	26.7	

STUDY II DATA

		Training		
		Treatment A	Treatment B	Row Mean
Control (Intelligence)	Level I (Low)	10.4	20.6	15.5
	Level II (High)	30.6	40.4	35.5
	Column mean	20.5	30.5	

Simple Factorial Design



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.
- 2 × 2 × 2 complex factorial design which will contain a total of eight cells as shown below

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

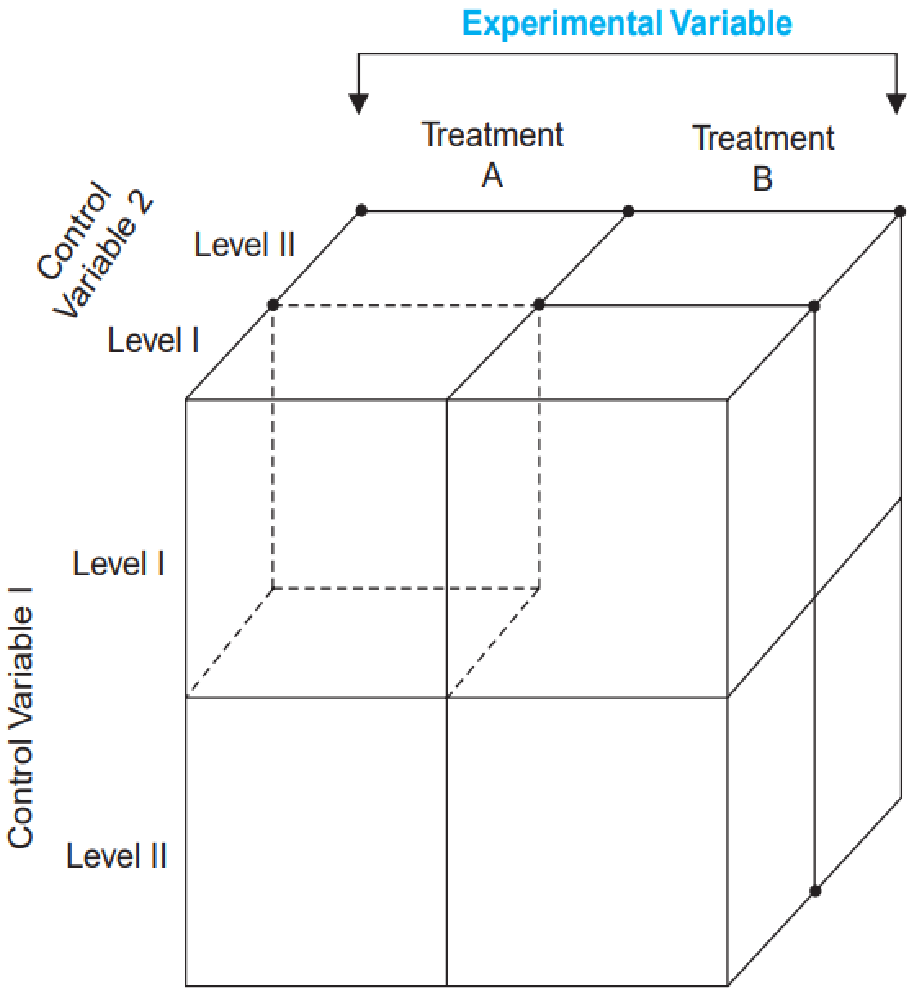
Complex Factorial Design

- 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.
- 2 × 2 × 2 complex factorial design which will contain a total of eight cells as shown below

2 × 2 × 2 COMPLEX FACTORIAL DESIGN

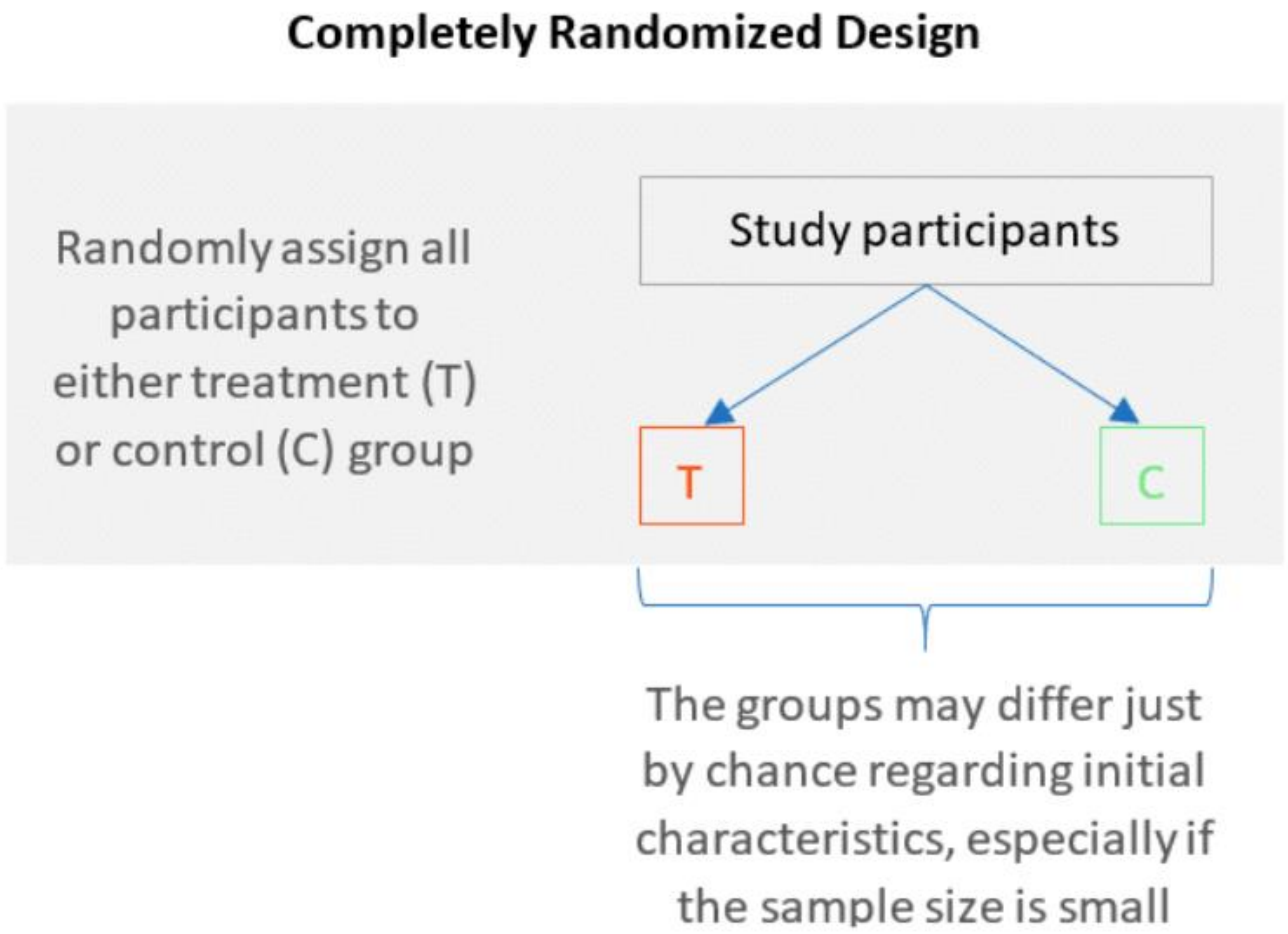
		Experimental Variable			
		Treatment A		Treatment B	
		Control Variable 2 Level I	Control Variable 2 Level II	Control Variable 2 Level I	Control Variable 2 Level II
Control Variable 1	Level I	Cell 1	Cell 3	Cell 5	Cell 7
	Level II	Cell 2	Cell 4	Cell 6	Cell 8

Complex Factorial Design



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

Completely Randomised Design



Completely Randomised Design

- It is commonly called as CRD.
- CRD is a statistical experimental design where the treatments are assigned completely at random so that each treatment unit has the same chance of receiving any one treatment.
- In CRD, any difference among experimental units receiving the same treatment is considered as an experimental error.
- CRD is applicable only when the experimental material is homogenous (Example: Homogenous soil condition in the field).
- Usually in the field, the soil will be HETEROGENOUS.
- Thus, CRD is not a preferable method in field experiments.
- CRD is generally applicable to the lab experimental conditions.
- In labs, the environmental conditions can be easily controlled.
- The concept of 'Local-control' is not used in CRD.

Completely Randomised Design

- Involves only two principles viz.,

✓ **Principle of replication**

- Even unequal replications can also work in this design.
  - Such a design is generally used when experimental areas are homogeneous.

✓ **Principle of randomization**

- Here subjects are randomly assigned to experimental treatments (or vice-versa).
  - For instance, if we have 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.

Disadvantages of factorial design

➢ The disadvantages of the factorial experiments are:

- It is disadvantageous because the execution of the experiment and the statistical analysis becomes more complex when several treatments combinations or factors are involved simultaneously.
- It is also disadvantageous in cases where may not be interested in certain treatment combinations but we are forced to include them in the experiment. This will lead to wastage of time and also the experimental material.
- In factorial experiments, the number of treatment combinations will increase if the factors are increased. This will also lead to the increase in block size, which in turn will increase the heterogenicity in the experimental material. Because of this it will lead to the increased experimental error and will decrease the precision in the experiment. Appropriate block size must be maintained.

Completely Randomised Design

Scenario: Testing a New Medication for Headaches
Imagine a pharmaceutical company is testing a new medication designed to relieve headaches. They want to know if this medication is effective, so they set up an experiment with two groups of participants: an experimental group and a control group.

1.Experimental Group:

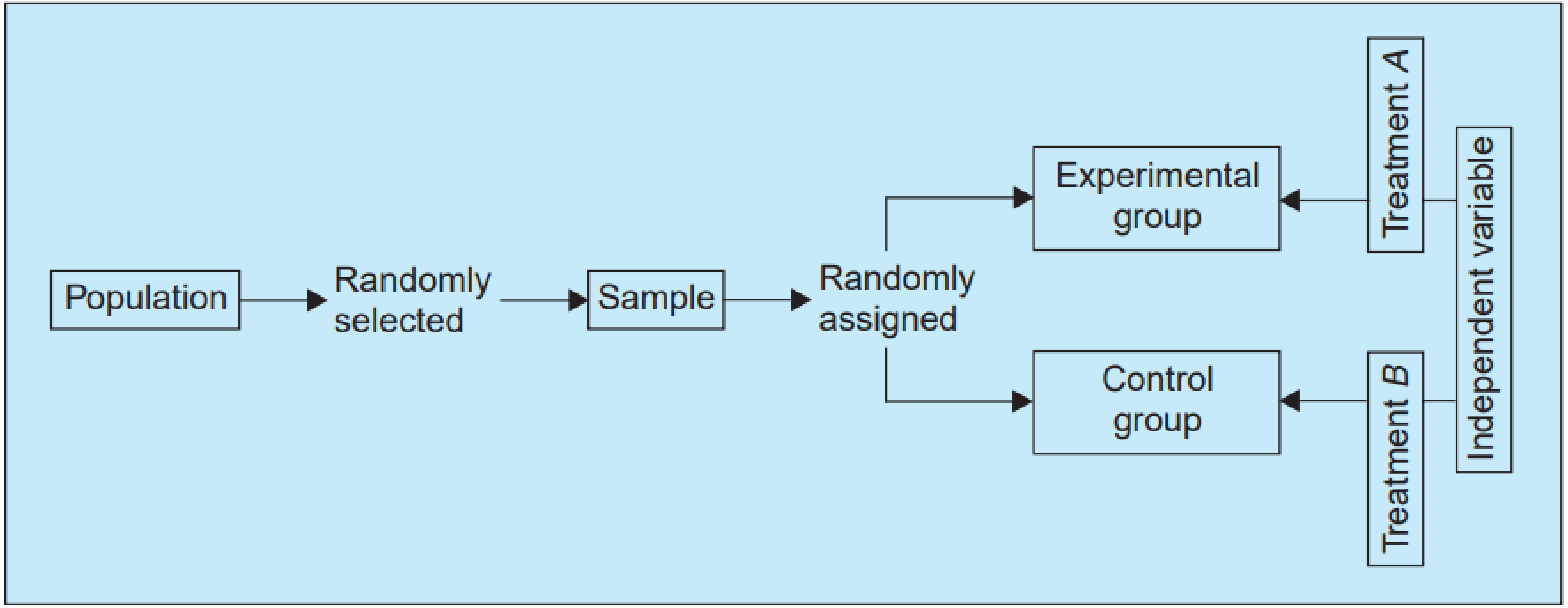
1. This group receives the new headache medication.
2. Participants in this group are the ones exposed to the "independent variable" (the medication) because the researchers are testing the medication’s effectiveness.
3. By giving the experimental group the medication, researchers can observe any changes in headache symptoms that might be caused by the medication.

2.Control Group:

1. The control group receives a placebo (a pill with no medicinal effects), rather than the real medication.
2. The purpose of the control group is to act as a baseline, helping researchers determine if any effects seen in the experimental group are actually due to the medication itself rather than other factors.
3. Because they’re not receiving the actual medication, any changes in headache symptoms in this group are likely due to psychological factors, the natural course of headaches, or other unrelated influences.

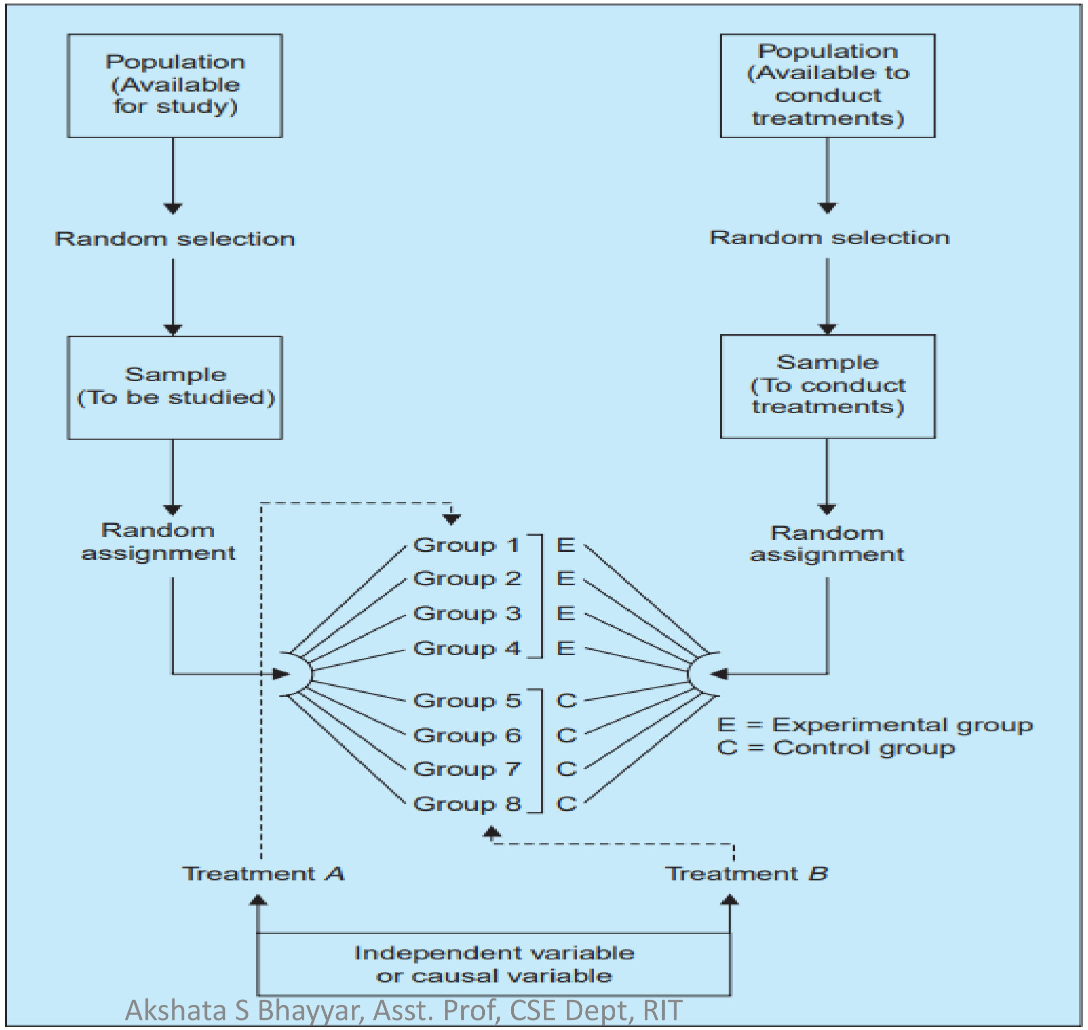
Completely Randomised Design

- (i) Two-group simple randomized design:



Completely Randomised Design

- (ii) Random replications design:



Advantages of CRD

- Ø CRD is easy to understand and calculate the variance.
- Ø The number of replications can vary from treatment to treatment.
- Ø CRD has high flexibility and thus any number of treatments can be used.
- Ø Simple statistical analysis is required in the analysis of CRD.
- Ø CRD provides maximum number of degree of freedom.

Disadvantages of CRD

- Ø CRD can be applied only to homogenous experiments.
- Ø The principle of 'Local-control' is not used in CRD.

Randomized block design (R.B. design)

- Here subjects are first divided into groups, known as blocks, such that within each group the subjects are relatively homogeneous in respect to some selected variable (Eg. Conference tracks)
- The variable selected for grouping the subjects is believed to be related to the measures to be obtained concerning the dependent variable.
- The number of subjects in a given block would be equal to the number of treatments and one subject in each block would be randomly assigned to each treatment.
- 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.

Randomized block design (R.B. design)

Example: Testing Fertilizers on Crop Yield
Imagine a research team that wants to test the effectiveness of three different fertilizers (A, B, and C) on crop yield. However, they know that **soil quality** (a factor beyond their control) varies across the field and may affect the results. To account for this, they decided to use a Randomized Block Design.

Steps in the RBD setup:

1. **Identify Blocks Based on a Factor**:
   1. In this example, soil quality is the blocking factor.
   2. The researchers divide the field into blocks based on soil quality—e.g., **Block 1: high-quality soil**, **Block 2: medium-quality soil**, and **Block 3: low-quality soil**.
2. **Randomly Assign Treatments Within Each Block**:
   1. Within each block, the researchers randomly assign the three fertilizers (A, B, and C) to specific plots of land.
   2. For example, in **Block 1 (high-quality soil)**, Fertilizer A, B, and C are each applied to one of the plots, and the assignment is random to avoid bias.
3. **Conduct the Experiment and Collect Data**:
   1. After applying the fertilizers and allowing time for growth, the researchers measure the crop yield in each plot.
4. **Analyze Results by Block**:
   1. By comparing the yields within each block, researchers can assess the effect of each fertilizer while minimizing the impact of soil quality differences.

Randomized block design (R.B. design)

	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

Advantages of RBD

- Ø RBD is more efficient and accurate when compared to CRD.
- Ø Chance of error in RBD is comparatively less.
- Ø Flexibility is also very high in RBD and thus any number of treatments and any number of replications can be used.
- Ø Statistical analysis is relatively simple and easy.
- Ø Statistical analysis simple when one value is missing.
- Ø Errors of any treatment can be isolated.

Disadvantages of RBD

- Ø RBD is not advised for very large number of treatments.
- Ø If the heterogeneity of the plot is very high, RBD cannot be applied. When the number of treatments is very large then the size of each block will be increased so that there may be heterogeneous blocks within.
- Ø With large number of treatments, the possibility of experimental errors will be high.

Latin Square Design

- It is an experimental design very frequently used in agricultural research.
  - For instance, an experiment has to be made through which the effects of five different varieties of fertilizers on the yield of a certain crop, say wheat, it to be judged.
  - In such a case the varying fertility of the soil in different blocks in which the experiment has to be performed must be taken into consideration; otherwise, the results obtained may not be very dependable because the output happens to be the effect not only of fertilizers, but it may also be the effect of fertility of soil.
  - Similarly, there may be impact of varying seeds on the yield.
- To overcome such difficulties, the L.S. design is used when there are two major extraneous factors such as the varying soil fertility and varying seeds.

Latin Square Design

FERTILITY LEVEL

Seeds differences

	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

Latin Square Design

Merits and Demerits

- It enables differences in fertility grants in the field to be eliminated in comparison to the effects of different varieties of fertilizers on the yield of the crop.

  Limitation:

  ✓it is that although each row and each column represents equally all fertilizer varieties, there may be considerable differences in the row and column means both up and across the field.

  ✓So L.S. design must assume that there is no interaction between treatments and blocking factors.

  ✓This defect can, however, be removed by taking the means of rows and columns equal to the field mean by adjusting the results.

- **Another limitation** of this design is that it requires a number of rows, columns, and treatments to be equal.

  ✓This reduces the utility of this design.

  ✓In the case of (2 × 2) L.S. design, there are no degrees of freedom available for the mean square error and hence the design cannot be used.

  ✓If treatments are 10 or more, then each row and each column will be larger in size so that rows and columns may not be homogeneous.

  ✓Therefore, L.S. design of orders (5 × 5) to (9 × 9) are generally used.

Thank You