Unit-3

RESEARCH DESIGN AND ANALYSIS

RESEARCH DESIGN

The task of defining the research problem is the preparation of the design 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 research design.

"A research design is the arrangement of conditions for collection and analysis of data in a manner that aims to combine relevance to the research purpose with economy in procedure."

In fact, the research design is the conceptual structure within which research is conducted; it constitutes the blueprint for the collection, measurement and analysis of data. As such the design includes an outline of what the researcher will do from writing the hypothesis and its operational implications to the final analysis of data.

A research design includes

(i) What is the study about?

(ii) Why is the study being made?

(iii) Where will the study be carried out?

(iv) What type of data is required?

(v) Where can the required data be found?

(vi) What periods of time will the study include?

(vii) What will be the sample design?

(viii) What techniques of data collection will be used?

(ix) How will the data be analysed?

(x) In what style will the report be prepared?

Keeping in view the above stated design decisions, the research design can be split into the following parts or groups of activities.

(a) the sampling design which deals with the method of selecting items to be observed for the given study.

(b) the observational design which relates to the conditions under which the observations are to be made;

(c) the statistical design which concerns with the question of how many items are to be observed and how the information and data gathered are to be analysed;

(d) the operational design which deals with the techniques by which the procedures specified in the sampling, statistical and observational designs can be carried out.

In brief, research design must, at least, contain

(i) a clear statement of the research problem;

(ii) procedures and techniques to be used for gathering information

(iii) the population to be studied and

(iv) methods to be used in processing and analysing data.

NEED FOR RESEARCH DESIGN

Research design is needed because it facilitates the smooth sailing of the various research operations, thereby making research as efficient as possible yielding maximal information with minimal expenditure of effort, time and money.

we need a research design or a plan in advance of data collection and analysis for our research project. 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.

Without proper research design many researches do not serve the purpose for which they are undertaken. In fact, they may even give misleading conclusions.

The research design helps the researcher to organize his ideas in a form whereby it will be possible for him to look for flaws and inadequacies.

DIFFERENT RESEARCH DESIGNS

The research designs can be categorized 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.

1. Research design in case of exploratory research studies

The 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. For these studies the research design should be flexible. Generally, the following three methods are used.

(a) the survey of concerning literature

(b) the experience survey and

(c) the analysis of 'insight-stimulating' examples.

The **survey of concerning literature** happens to be the most simple and fruitful method of formulating precisely the research problem or developing hypothesis. Hypotheses stated by earlier researchers may be reviewed and their usefulness be evaluated as a basis for further research.

Experience survey means the survey of people who have had practical experience with the problem to be studied. The object of such a survey is to obtain insight into the relationships between variables and new ideas relating to the research problem. For such a survey people who are competent and can contribute new ideas may be carefully selected as respondents to ensure a representation of different types of experience. The respondents so selected may then be interviewed by the investigator. The researcher must prepare an interview schedule for the systematic questioning of informants. But the interview must ensure flexibility in the sense that the respondents should be allowed to raise issues and questions which the investigator.

Analysis of 'insight-stimulating' examples is also a fruitful method for suggesting hypotheses for research. It is particularly suitable in areas where there is little experience to serve as a guide. This method consists of the intensive study of selected instances of the phenomenon in which one is interested. For this purpose, the existing records, if any, may be examined, the unstructured interviewing may take place.

In an exploratory of formulative research study which merely leads to insights or hypotheses, whatever method or research design outlined above is adopted, the only thing essential is that it must continue to remain flexible so that many different facets of a problem may be considered as and when they arise and come to the notice of the researcher.

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.

Studies concerned with specific predictions, with narration of facts and characteristics concerning individual, group or situation are all examples of descriptive research studies.

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.

From research design point of view both studies share common requirements and as such we may group together these two types of research studies. In descriptive as well as in diagnostic studies, the researcher must be able to define clearly, what he wants to measure and must find adequate methods for measuring it along with a clear-cut definition of 'population' he wants to study.

These studies must be rigid and not flexible and must focuses 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.

In a descriptive/diagnostic study the first step is to specify the objectives with sufficient precision to ensure that the data collected are relevant. If this is not done carefully, the study may not provide the desired information.

Then comes the question of selecting the methods by which the data are to be obtained. In other words, techniques for collecting the information must be devised. Several methods (viz., observation, questionnaires, interviewing, examination of records, etc.).

In these studies, the researcher will select the sample. The data collected must be processed and analysed. This includes steps like coding the interview replies, observations, etc.; tabulating the data; and performing several statistical computations. Reporting the findings involve communicating the findings to others.

The Difference between Exploratory and Descriptive/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.

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

Hypothesis-testing research studies (generally known as experimental 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 Designs

Professor R A Fisher has enumerated three principles of experimental designs:

(1) the Principle of Replication;

(2) the Principle of Randomization; and the

(3) Principle of Local Control.

(1)Principle of Replication:

According to this 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.

For example, suppose we are to examine the effect of 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. We can then compare the yield of the two parts and draw conclusion on that basis. But if we are to apply the principle of replication to this experiment, then we first divide the field into several parts, grow one variety in half of these parts and the other variety in the remaining parts. We can then collect the data of yield of the two varieties and draw conclusion by comparing the same. The result so obtained will be more reliable in comparison to the conclusion we draw without applying the principle of replication. The entire experiment can even be repeated several times for better results.

(2)The Principle of Randomization

It 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."

For instance, if we grow one variety of rice, say, in the first half of the parts of a field and the other variety is grown in the other half, 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. In such a situation, we may assign the variety of rice to be grown in different parts of the field on the basis of some random sampling technique i.e., we may apply randomization principle and protect ourselves against the effects of the extraneous factors (soil fertility differences in the given case).

(3)The Principle of Local Control

Under this principle of experimental 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. This means that we should plan the experiment in a manner that we can 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) and experimental error.

In other words, according to the principle of local control, we first divide the field into several homogeneous parts, known as blocks, and then each such block is divided into parts equal to the number of treatments. Then the treatments are randomly assigned to these parts of a block. Dividing the field into several homogenous parts is known as 'blocking'.

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

Developing a Research Plan

After identifying and defining the problem as also accomplishing the relating task, researcher must arrange his ideas in order and write them in the form of an experimental plan or what can be described as 'Research Plan'. This is essential specially for new researcher because of the following:

(a) It helps him to organize his ideas in a form whereby it will be possible for him to look for flaws and inadequacies, if any.

(b) It provides an inventory of what must be done and which materials have to be collected as a preliminary step.

(c) It is a document that can be given to others for comment.

Research plan must contain the following items.

1. Research objective should be clearly stated in a line or two which tells exactly what it is that the researcher expects to do.

2. The problem to be studied by researcher must be explicitly stated so that one may know what information is to be obtained for solving the problem.

3. Each major concept which researcher wants to measure should be defined in operational terms in context of the research project.

4. The plan should contain the method to be used in solving the problem. An overall description of the approach to be adopted is usually given and assumptions, if any, of the concerning method to be used are clearly mentioned in the research plan.

5. The plan must also state the details of the techniques to be adopted. For instance, if interview method is to be used, an account of the nature of the contemplated interview procedure should be given. Similarly, if tests are to be given, the conditions under which they are to be administered should be specified along with the nature of instruments to be used. If public records are to be consulted as sources of data, the fact should be recorded in the research plan. Procedure for quantifying data should also be written out in all details.

6. A clear mention of the population to be studied should be made. If the study happens to be sample based, the research plan should state the sampling plan i.e., how the sample is to be identified. The method of identifying the sample should be such that generalisation from the sample to the original population is feasible.

7. The plan must also contain the methods to be used in processing the data. Statistical and other methods to be used must be indicated in the plan. Such methods should not be left until the data have been collected. This part of the plan may be reviewed by experts in the field, for they can often suggest changes that result in substantial saving of time and effort.

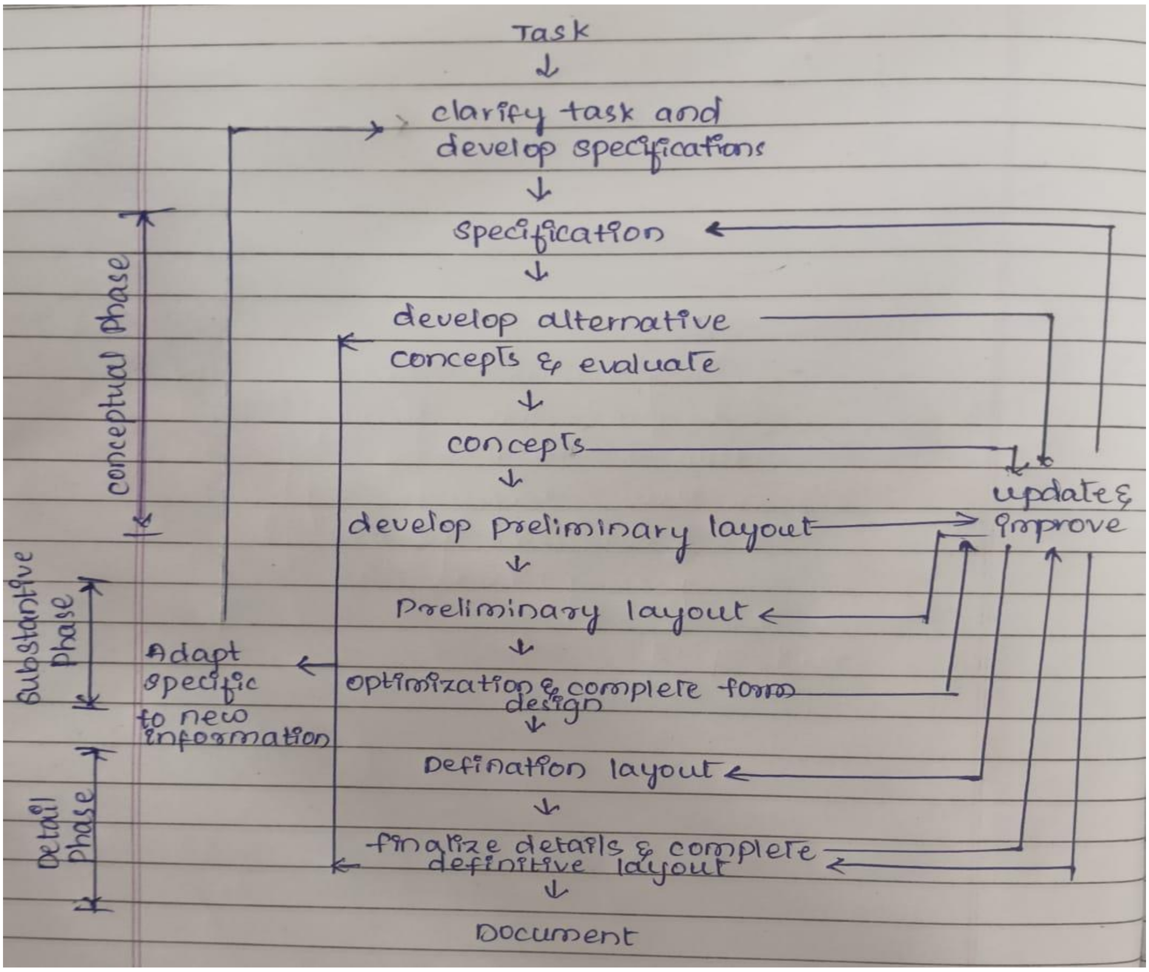
8. Results of pilot test, if any, should be reported. Time and cost budgets for the research project should also be prepared and laid down in the plan itself.

Design of experimental set up

Experimental set up is developed to meet the specific need of the researcher using experimental set up the researcher finds the effect of independent variable(s) on the dependent variable. The experimental set-up may vary depending upon the objective of the research study.

For the development of experimental set-up one can follow the approach recommended for a new product design. Experimental set-up is designed to meet the specific needs of the researcher. Using an appropriate experimental set-up, the researcher finds the effect of independent variable(s) on the dependent variable. It may vary depending upon the objective of the research study. Designing an experimental set-up or developing a new product for a specific purpose are similar.

To develop an experimental set-up, follow the approaches recommended for a new product design, as shown in the flowchart below.



The whole processes are divided into 3 major phares

- ➢ Conceptual
- ➢ Substantive and
- ➢ detail design.

During conceptual phase, think all the different ways to accomplish the basic task the product has to perform, evaluate each approach and select those which are feasible ones.

In the substantive phase, think of physical device which can be actual for function analysed for and cost. Each selected approach meant be carefully analysed, improved and adapted to our specific needs and then critically evaluated against performance and cost Specifications. Then add new concepts to initial list. At the end of this phase superior alternative design is selected.

The analogy of product design can be adopted for preparing an experimental set-up needed by a researcher for their study. During the conceptual phase, avoid thinking about physical hardware and concentrate on functional requirements. The steps that may help move from specifications to one or more concepts include:

- Identifying the essential functions from the list of considerations.
- Establishing function structures (overall and sub-function).
- Searching for solution principles to fulfil the overall function.
- Selection of suitable combination.
- Firm up concept variants.

The approaches that help to search for solution principles to fulfil subfunctions are:

(i) Literature search

(ii) Known technical systems and analogies

(iii) Previous experimental studies

(iv) Design catalogues

(v) Discussion with colleagues and experts, etc.

In the substantive or embodiment phase of experimental set-up design, the researcher should concentrate on how to convert conceptual set-ups into physical or simulated structures. Here, the main emphasis is on the identification and assessment of equipment, instruments, sources of supply, their methods of functioning, precision, cost, etc., for physical experiments. For simulation studies, the availability and efficiency of software, as well as its limitations, should be considered in respect of various alternative concepts developed in the first phase. More weightage should be given to technical functioning.

The final phase is a detailed planning phase. Hence, each alternative experimental set-up has to be thoroughly evaluated, taking into account the study objective, expected outcome, feasibility, cost involved, and time frame required. After evaluating each alternative, the most feasible one from all considerations has to be selected for adoption in the study

Use of standards and codes

Codes and standards are an important source of documents to help the researcher in the design of experimental set-ups.

Standards cover instrument accuracy, test apparatus, testing methods and environment. They specify the number of specimens to be taken and the statistical tools to be used, which are useful in sampling.

Many developed and developing countries have their national standards organizations. These organizations ensure the coordination and preparation of standards and codes relating to various fields. USA has the American National standards Institute (ANSI), British Standards Institution (BSI) in the UK, and Deutches Institute for Normung (DIN) in Germany are examples of national standards organizations. Internationally, the International Organization for Standardization (ISO) examines the standards prepared by member countries and evolves common international standards.

In addition to the above, various professional associations are also involved in the preparation of specific standards and codes on various aspects. For example, the American Society for Testing and Materials (ASTM).

Important Experimental Designs

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, viz., informal experimental designs and formal experimental designs. Informal experimental designs are those designs that normally use a less sophisticated form of analysis based on differences in magnitudes, whereas formal experimental designs offer relatively more control and use precise statistical procedures for analysis.

(a) **Informal experimental designs:**

(i) Before-and-after without control design.

(ii) After-only with control design.

(iii) Before-and-after with control design.

(b) **Formal experimental designs:**

(i) Completely randomized design (C.R. Design).

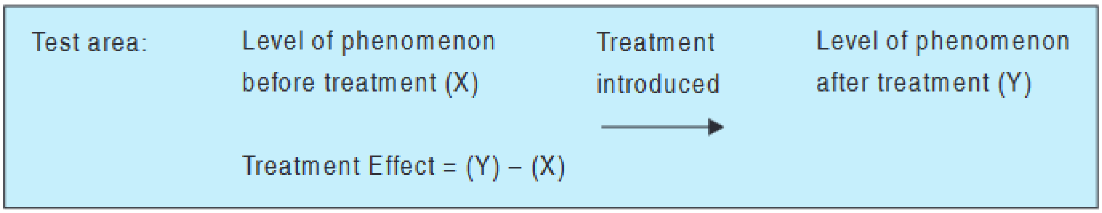
(ii) Randomized block design (R.B. Design).

(iii) Latin square design (L.S. Design).

(iv) Factorial designs.

1. Before-and-after without control design:

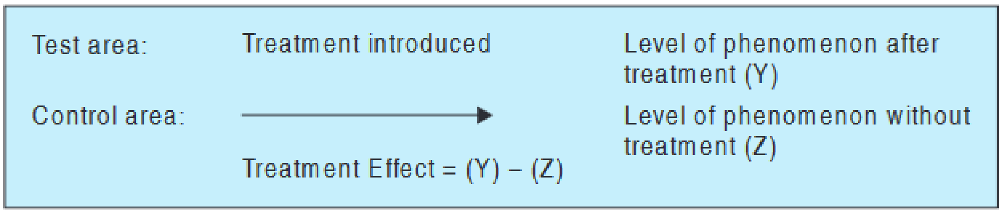
In such a design a single test group or area is selected and the dependent variable is measured before the introduction of the treatment. The treatment is then introduced and the dependent variable is measured again after the treatment has been introduced. The effect of the treatment would be equal to the level of the phenomenon after the treatment minus the level of the phenomenon before the treatment. The design can be represented as



2. After-only with control design:

In this design two groups or areas (test area and control area) are selected and the treatment is introduced into the test area only. The dependent variable is then measured in both the areas at the same time. Treatment impact is assessed by subtracting the value of the dependent variable in the control area from its value in the test area.

The basic assumption in such a design is that the two areas are identical with respect to their behaviour towards the phenomenon considered. If this assumption is not true, there is the possibility of extraneous variation entering into the treatment effect. The design can be represented as



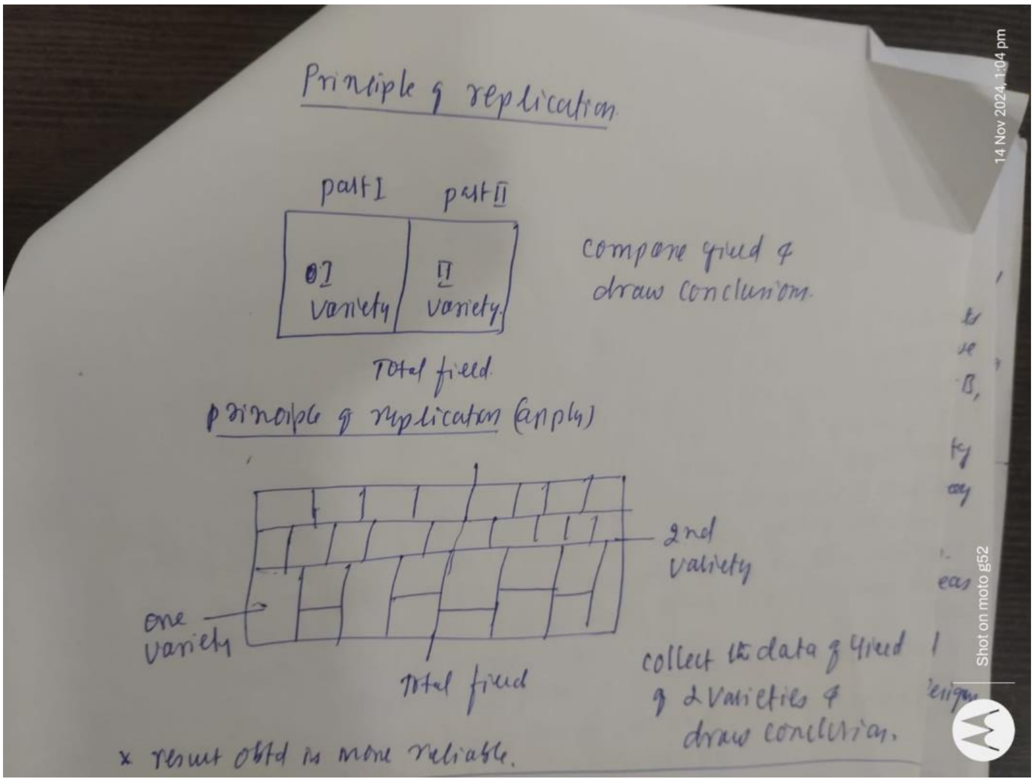
3. Before-and-after with control design:

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 treatment effect is determined by subtracting the change in the dependent variable in the control area from the change in the dependent variable in test area. The design can be represented as

	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)

This design is superior to the above two designs for the simple reason that it avoids extraneous variation resulting both from the passage of time and from non-comparability of the test and control areas. due to lack of historical data, we prefer design 1 or design2.

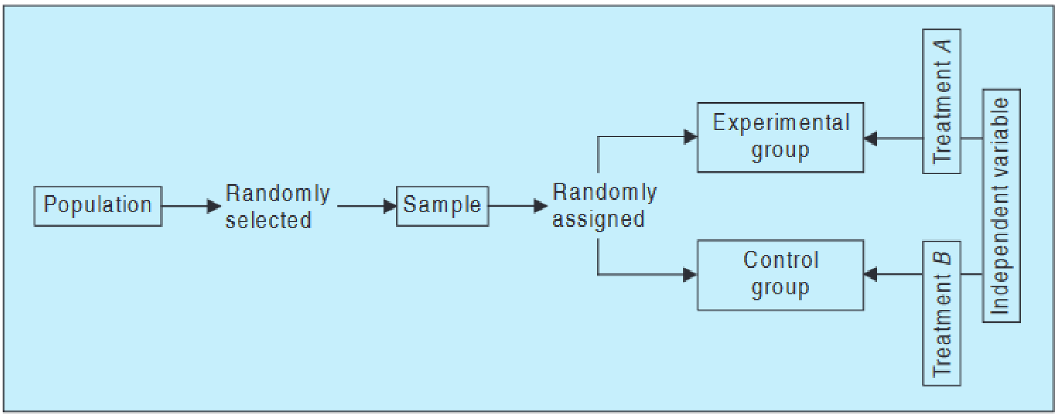


Formal experimental designs Completely randomized design (C.R. design): Involves only two principles viz., the principle of replication and the principle of randomization of experimental designs. It is the simplest possible design and its procedure of analysis is also easier. The essential characteristic of the design is that subjects are randomly assigned to experimental treatments (or vice-versa). 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. One-way analysis of variance is used to analyse such a design. Even unequal replications can also work in this design. It provides maximum number of degrees of freedom to the error. Such a design is generally used when experimental areas happen to be homogeneous.

There are two forms of completely randomized designs

- Two-group simple randomized design
- Random replications design

(i) **Two-group simple randomized design**: In a two-group simple randomized design, first of all the population is defined and then from the population a sample is selected randomly. Further, requirement of this design is that items, 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). Thus, this design yields two groups as representatives of the population. In a diagram form this design can be shown in this way Since in the sample randomized design the elements constituting the sample are randomly drawn from the same population and randomly assigned to the experimental and control groups, it becomes possible to draw conclusions on the basis of samples applicable for the population.



The two groups (experimental and control groups) of such a design are given different treatments of the independent variable.

This design of experiment is quite common in research studies concerning behavioural sciences.

The merit of such a design is that it is simple and randomizes the differences among the sample items.

But the limitation of it is that the individual differences among those conducting the treatments are not eliminated, i.e., it does not control the extraneous variable.

[Suppose the researcher wants to compare two groups of students who have been randomly selected and randomly assigned. Two different treatments viz., the usual training and the specialised training are being given to the two groups. The researcher hypothesises greater gains for the group receiving specialised training. To determine this, he tests each group before and after the training, and then compares the amount of gain for the two groups to accept or reject his hypothesis. This is an illustration of the two-groups randomized design, where individual differences among students are being randomized.]

(ii) 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’.

Random replication design serves two purposes;

- It provides controls for the differential effects of the extraneous independent variables.
- It randomizes any individual differences among those conducting the treatments.

[From the diagram it is clear that there are two populations in the replication design. The sample is taken randomly from the population available for study and is randomly assigned to, say, four experimental and four control groups. Similarly, sample is taken randomly from the population available to conduct experiments (because of the eight groups eight such individuals be selected) and the eight individuals so selected should be randomly assigned to the eight groups. Generally, equal number of items are put in each group so that the size of the group is not likely to affect the result of the study. Variables relating to both population characteristics are assumed to be randomly distributed among the two groups.].

Randomized block design (R.B. design):

It is an improvement over the C.R. in this design all these basic principles of experimental design are used.

In the R.B. design, 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. The variable selected for grouping the subjects is one that is believed to be related to the measures to be obtained in respect of 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.

The R.B. design is analysed by the two-way analysis of variance technique.

[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.]

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.

	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

Latin square design (L.S. design)

This type of experimental design very frequently used in agricultural research. The conditions under which agricultural investigations are carried out are different from those in other studies for nature plays an important role in agriculture.

For example, 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 majors' extraneous factors such as the varying soil fertility and varying seeds.

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

The two blocking factors may be represented through rows and columns. Fig shows diagrammatized from L.S design The following is a diagrammatic form of such a design in respect of, say, five types of fertilizers, viz., A, B, C, D and E and the two-blocking factor viz., the varying soil fertility and the varying seeds.

Factorial designs:

Factorial designs are used in experiments where the effects of varying more than one factor is to be determined. They are especially 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 and
(ii) complex factorial designs. We take them separately

(i) Simple factorial designs: 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'.

(ii) Complex factorial designs: 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.

Factorial designs are used mainly because of the two advantages.

(i) They provide equivalent Accuracy with less labour and as such are a source of economy.

(ii) They permit various other comparisons of interest.

Testing of Hypothesis

Ordinarily, hypothesis is simply means a more assumption or supposition to be proved or disproved. But for a researcher hypothesis is a formal question that he intents to resolve. Thus a hypothesis may be defined as a proposition or a set of proposition set forth as an explanation for the occurrence of some specified group of phenomena either asserted merely as a conjecture to guide some investigation or accepted as highly probable in the light of established facts.

(OR)

Research hypothesis is a predictive statement, capable of being tested by scientific methods, that relates an independent variables to some dependent variable.

For example, "students who receive counselling will show a greater increase in creativity than students not receiving counselling".

(OR)

"The automobile A is performing as well as automobile B"

Characteristics of hypothesis

Hypothesis must posses the following characteristics

i) Hypothesis should be clear and precise.
ii) Hypothesis being should be capable of being tested.
iii) Hypothesis should state relationship between variables.
iv) Hypothesis should be limited in scope and must be specific.
v) Hypothesis should be stated as far as possible in most simple terms so that the same is easily understandable by all concerned.
vi) Hypothesis should be consistent with most known facts i.e. it must be consistent with a substantial body of established facts.
vii) Hypothesis must explain the facts that gave rise to the need for explanation

Basic concepts concerning testing of hypothesis

1. **Null hypothesis and Alternate hypothesis**.

If we are to compare method A with method B about the superiority and if we proceed on the assumption that both the methods are equally good. Then this assumption is termed as the null hypothesis.

As against this, we may think that method A to superior or method B is inferior, we are then stating what is termed as alternative hypothesis. The null hypothesis is generally symbolized has Ho and the alternative hypothesis as H_1.

Alternative hypothesis is usually the one which one wishes to prove and the null hypothesis is the one which one wishes to disprove. Thus, null hypothesis represents the hypothesis we are trying to reject and alternative hypothesis represents all other possibilities.

2.Type I and Type II errors

In its context of testing of hypothesis basically there are 2 types of errors. We may reject Ho when Ho is true and we may accept Ho when infact Ho is not true. The former in known as type I error and the later as Type II error.

In other words Type I error means rejection of hypothesis which should have been accepted and type II error means accepting the hypothesis which showed how been rejected.

3. Level of significance.

It is the probability of Type I error .This is very important concept in the context of hypothesis testing. It is always some percentage (usually 5%) which should chosen with great care, though and reason 5% level of significance means that the researcher is willing to take as much as a 5% risk & rejecting the null hypothesis when it happens to be true. Thus significance level is Its maximum value of the probability of rejecting Ho when it is true and is usually determined in advance before testing the hypothesis.

4.Two-Tailed and one-Tailed tests

We test 3 types of hypothesis

i) Ho: M=Mo Against H_1; M $\neq$ Mo.
ii) Ho: M=Mo Against Ha: M > Mo or Ho: M $\leq$ Mo Against H_1: M>Mo.
iii) Ho: M=Mo Against Ha: M < Mo or Ho: M $\geq$ Mo Against H1 : M < Mo.

on the sign in alternative hypothesis ($\neq$,> or <) we have 3 different tests. Then we have '$\neq$' sign in alternative hypothesis, we have two-tailed tests, when we have '>' sign in alternative hypothesis we have right tailed test and for '<'sign is alternative hypothesis, we have left tailed test.

Procedure for hypothesis testing

In hypothesis testing the main question is whether to reject the null hypothesis or not to reject the null hypothesis?

Procedure for hypothesis testing refers to all these steps that we undertake for making its choice the between two actions I.e rejection and acceptance of a null hypothesis. The Various steps are

(i) Setting up the hypothesis

This consists & making a formal statement of the null hypothesis (Ho) and also the alternative hypothesis (H1). This means that hypothesis should be clearly stated, considering the Nature of the problem.

For example, Mr. Mohan of the Civil Engg department wants to test the load bearing capacity of an old bridge which must be more than 10 tons, in that he can state his hypothesis as

Null hypothesis Ho: M=10 tons.

Alternate hypothesis H1: M> 10 tons.

(ii) selecting a significance level.

Hypothesis are tested on a predetermined level of significance and as such the name should be specified. Generally, in practice, 5% level or 1% level is accepted. The factors that affect the level of significance are

- The magnitude of the difference between sample means
- The size of the samples
- The variability of measurements within samples
- Whether the hypothesis is directional or non-directional

(iii)Test statistic

Hypothesis tests are conducted for mean, proportion and variance. The value of test statistic is obtained using the sample observations selected by the researcher and the hypothetical parametric Value stated under null hypothesis.

(iv) Critical value: using the distribution of test statistic, level of significance , α and the type of test (two-tailed, right tailed or left tailed), we obtain the critical Value.

(v)Decision: comparing the value of test statistic and critical value, we make the decision about rejecting or not rejecting the null hypothesis.

We rejecting null hypothesis when

a) value of test statistic < lower critical value(two-tailed)

b) Value of test statistic > critical value, in case of right -tailed test

c) values test static < critical value, in case & left -tailed test.

[Test statistic - It is calculated from sample observations.

Critical value- It is obtained from sampling distribution of test statistic and prefixed level of significance.]

OVERVIEW OF MULTIVARIATE ANALYSIS

Definition :

Multivariate analysis is a powerful statistical tool that enables researchers to investigate and understand complex data sets by considering multiple variables simultaneously. Unlike univariate analysis, which examines one variable at a time, multivariate analysis explores the relationships and interactions among several variables.

Another author, defines Multivariate analysis as "those statistical techniques which focus upon, and bring out in bold relief, the structure of simultaneous relationships among three or more phenomena.

The Multivariate analysis is a statistical technique that examines the relationship between multiple variables at once (3 or more variables)

Purpose: The main goals of multivariate analysis include:

- Understanding complex relationships among variables.
- Identifying patterns and trends in data.
- Reducing data dimensionality while retaining essential information.
- Making predictions based on multiple inputs.

Common Techniques

1. **Multiple Regression Analysis**:

Examines the relationship between one dependent variable and multiple independent variables to predict outcomes. Ex: researchers might investigate how various factors, such as age, income, education level, and lifestyle choices, impact individuals' health outcomes (the dependent variable).

2. **Factor Analysis**:

Identifies underlying factors that explain the correlations among multiple observed variables, useful for data reduction.

Ex: Researchers might collect survey data on responses to questions about consumer preferences for various smartphone features such as battery life, camera quality, screen size, brand reputation, and price.

Through factor analysis, the researchers could identify: Performance: includes battery life and processing speed.

Camera Quality: encompasses camera resolution and features.
Value: combines brand reputation and price sensitivity.

3. **Principal Component Analysis (PCA)**:

Transforms correlated variables into a set of uncorrelated variables (principal components) to simplify data analysis. These variables represent the directions of maximum variance in the data.

Ex: In analyzing air quality data, researchers collect data on various air pollutants, including levels of carbon monoxide, nitrogen dioxide, sulfur dioxide, ozone etc. across multiple cities. Using PCA, the researchers can reduce the complexity of this data as follows:

Component 1: Represents overall pollution levels (combining effects of all pollutants).
Component 2: Distinguishes between traffic-related pollutants (e.g., carbon monoxide and nitrogen dioxide) and industrial pollutants (e.g., sulfur dioxide).

4. **Cluster Analysis**:

Groups similar observations or variables into clusters, helping to identify natural groupings within data.

Ex: A retail company collects data on various customer attributes, such as age, income, shopping frequency, and preferences for product categories (e.g., electronics, clothing, groceries). Using cluster analysis, the company can identify:

Cluster: 1- Younger customers with high incomes frequently buy electronics. Cluster: 2- Middle-aged customers with moderate incomes look for discounts. Cluster: 3- Young adults interested in clothing and accessories, shop regularly for latest trends.

5. MANOVA (Multivariate Analysis of Variance):

Tests differences in multiple dependent variables across groups based on independent variables.

Ex: A study is investigating the effects of three teaching methods (traditional, online, and hybrid) on student performance across multiple subjects, such as math, science, and language arts. The independent variable would be the teaching method, while the dependent variables would be the test scores in math, science, and language arts.

Using MANOVA, the researchers can determine if there are any statistically significant differences in student performance across the three teaching methods for the combined scores in all subjects.

Applications

- Market Research: Understanding consumer preferences and behavior.
- Healthcare: Analyzing patient data to identify risk factors for diseases.
- Social Sciences: Studying relationships between socioeconomic factors and outcomes.
- Finance: Portfolio management and risk assessment.