CPADA(C III)

CERTIFICATE PROGRAMME IN APPLIED DATA ANALYSIS

(CBCS in Accordance with NEP 2020)

STATISTICS IN RESEARCH



KRISHNA KANTA HANDIQUI STATE OPEN UNIVERSITY

Subject Experts

Professor Madhurjya Prasad Bezbaruah, Dept. of Economics, Gauhati University

Professor Nissar Ahmed Barua, Dept. of Economics, Gauhati University

Dr. Gautam Mazumdar, Dept. of Economics, Cotton University

Course Co-ordinators: **UTPAL DEKA,** KKHSOU

SLM Preparation Team

Units	Contributors
1	**Dr. Tarakeswar Chaudhury**, Retired Professor, Cotton College
2, 3	**Sunny Shreya Dutta,** Research Scholar, KKHSOU, Guwahati, Assam.
4	**Dr. Shaveta Kohli,** Department of Economics, Central University of Jammu
5	**Dr. Pranjal Sarma**, Department of Statistics L. C. B. College, Guwahati
6, 7	**Dr. Mitali Deka,** NERIM, Guwahati
8	**Dr. Bhaskar Sarmah,** KKHSOU; **Utpal Deka**, KKHSOU
9	**Dr. Bhaskar Sarmah,** KKHSOU
10,11	**Mrs. Poonam Rani,** UP Institute of Design, (Dr. APJ Abdul Kalam Technical University, Lucknow, Uttar Pradesh), Noida
12	**Dola Borkataki**, KKHSOU,**Shilpishikha Dowerah**, MSSV, Nagaon
13	**Dr. Mridusmita Duara**, KKHSOU; **Murchana Kaushik**, KKHSOU; **Shilpishikha Dowerah**, MSSV, Nagaon
14	**Dola Borkataki**, KKHSOU, **Dr. Mridusmita Duara**, KKHSOU

Editorial Team

Content	:	**Dr. Joydeep Baruah,** KKHSOU; **Utpal Deka,** KKHSOU; **Dr. Jagannath Ambagudia,** Tata Institute of Social Sciences, Guwahati; **Dr. Rajshree Bedamatta,** IIT, Guwahati; **Dr. Arunima Deka,** OKDISCD; **Dr. Parag Dutta,** KKHSOU; **Professor Tarakeswar Choudhary** (Retired), Cotton College, Guwahati
Structure, Format & Graphics	:	**Utpal Deka**, KKHSOU

August, 2023

© Krishna Kanta Handiqui State Open University

 This Self Learning Material (SLM) of the Krishna Kanta Handiqui State Open University is made available under a Creative Commons Attribution-Non Commercial-Share Alike 4.0 License (international): http://creativecommons.org/licenses/by-nc-sa/4.0/

Printed and published by Registrar on behalf of the Krishna Kanta Handiqui State Open University.

Headquarters: Patgaon, Rani Gate, Guwahati-781017

City Office: NH 37, Resham Nagar, Khanapara, Guwahati 781022; Web: www.kkhsou.in

The University acknowledges with thanks the financial support provided by the **Distance Education Bureau, UGC** for the preparation of this study material.

BACHELOR OF ARTS

Quantitative Methods for Economic Analysis - II

CONTENTS

		Pages
Unit 1:	**Basic Concepts in Sampling** Definition of Population, Parameter and Sampling: Random and Non-random Sampling Methods; Sampling with and without Replacement, Large Samples; Errors in Sampling Progress	**7-24**
Unit 2:	**Sampling Techniques- I** Random Sampling, Stratified Random Sampling, Systematic Sampling, Cluster Sampling	**25-37**
Unit 3:	**Sampling Techniques- II** Non-Random Sampling Methods: Convenience Samplin, Purposive Sampling, Quota Sampling, Snowball Sampling, Volunteer Sampling	**38-50**
Unit 4:	**Sampling Techniques- III** Meaning and Definition of Variables in Research; Various Types of Variables: Quantitative Variables, Qualitative Variables, Independent and Dependent Variables, Extraneous Variables, Background Variables, Intervening Variables, Suppressor Variables, Moderating Variables; Population or Universe of a Study: Finite and Infinite Population, Homogeneous and Heterogeneous Population, Existent and Hypothetical Population, Known and Unknown Population; Sample as a Concept: Characteristics of an Ideal / Good Sample, Population and Sampling Methods, Difference between Census and Sampling Methods	**51-69**
Unit 5:	**Characterisitcs of a Good Estimator** Estimators and Estimates; Types of Estimation; Criteria for a Good Estimator: Unbiasedness, Consistency, Efficiency	**70-81**
Unit 6:	**Testing of Hypothesis** Statistical Hypothesis: Null Hypothesis and Alternative Hypothesis; Errors in Hypothesis Testing, Level of Significance and Critical Region; One -Tailed and Two-Tailed Tests; General Procedure for Hypothesis Testing; Testing of Hypothesis in case of Large Samples: Hypothesis Testing for Single Population Mean, Hypothesis Testing for Single Population Proportion, Hypothesis Testing for Difference between Two Population Means, Hypothesis Testing for Difference between Two Population Means	**82-112**

Unit 7: Test of Significance **113-157**

The Chi-Square (χ^2) Distribution; Applications of χ^2 Test: χ^2 Test of Goodness of Fit, χ^2 Test for Independence of Attributes, χ^2 Test for a Specified Population Variance, χ^2 Test of Homogeneity; Testing of Hypothesis in case of Small Samples (t-Test): Characteristics of t-distribution, Applications of t-distribution, Hypothesis Testing for Single Population Mean, Hypothesis Testing for Difference between Two

UNIT 8: Basic Concepts of Research **158-175**

Research: Meaning and Definition; Nature of Research; Types and Importance of Research; Elements of Scientific Methods; Research Ethics; Significance of Statistics in Research

Unit 9: Types of Research **176-192**

Broad Types of Research; Types of Research based on its Purpose; Types of Research based on the Process; Types of Research based on the Outcome; Conceptual and Empirical Research

Unit 10: Review of Literature **193-204**

Review of Literature: Meaning and Concept: The 5 C's of a Literature Review; Need for Reviewing Literature; Sources of Literature

Unit 11: Research Problem **205-215**

Identification of the Research Problem; Selection of Research Problem: Contributing Factors; Technique Involved in Defining a Problem

Unit 12: Research Designs **216-225**

Meaning And Definition Of Research Design; Components Of Research Design: Title Of The Study, Statement Of The Problem, Review Of Previous Studies, Objectives Of The Study, Formulation Of Hypotheses, Methods Of Data Collection, Tools For Data Collection, Sampling; Need and importance of research design; Features of good research design

Unit 13: Types Of Research Design **226-248**

Exploratory Research Design; Descriptive Research Design; Experimental Research Design; Diagnostic Research Design, Survey and Case Study Methods; Difference Between Quantitative And Qualitative Research Designs

Unit 14: Writing Research Report **249-260**

Research Report: Meaning and Significance; Different Steps in writing the Report; Contents of a Research Report; Precautions in Writing Research Reports; Research Report Writing; Concept of Research Proposal and Research Abstract

Reference **261-262**

COURSE INTRODUCTION

This is the third course of the 'Certificate Programme in Applied Data Analysis'. It introduces us about the role of Statistics in Research. This Course has been divided into two fourteen units. The first four units of the course deals with sampling technique.The first unit will introduce you to the meaning of Sampling and types of sampling methods. After going through this unit, you will be able to define the term population, parameter and Statistics and know the random and non-random sampling methods. The second unit deliberates on the concept of random sampling, stratified random sampling, systematic sampling and cluster sampling. The third unit deals with the concept of non-random sampling methods like convenience sampling, purposive sampling, quota sampling, snowball sampling and volunteer sampling. The fourth unit of the course deals with the meaning, definition and types of variables in research.The fifth unit deliberates on the criteria for a good estimator like unbiasedness, consistency, efficiency. In the sixth unit we shall discuss about statistical hypothesis, errors in hypothesis testing and level significance.The seventh unit deliberates on the concept of test of significance. In this unit we will study about the Chi-Square Distribution and t-Distribution with its applications. From the eight unit onwards of this course, we will discuss about research and research design. In the eight unit we will discuss about the nature, importance and elements of Research.The ninth discuss about broad types of research like types of research based on its purpose, types of research based on the process, types of research based on the outcome, conceptual and empirical research.The tenth unit gives us a broad overview of review of literature. The eleventh unit deliberates on the research problem. In this unit we will discuss about identification of the research problem and selection of research problem. In the twelveth unit we shall discuss about the research design and its components. The thirteenth unit deals with the types of research design. The fourteenth unit, which is the last unit of the course deliberates on research report.Under this unit we shall study about the meaning, significance and steps of research report.

While going through a unit, you will notice some along-sode boxes, which have been included to help you know some of the difficult, unseen terms. Some 'ACTIVITY' (s) has been included to help you apply your own thoughts. Again, we have included some relevant concepts in 'LET US KNOW' along with the text. At the end of sections, you will get 'CHECK YOUR PROGRESS' questions. These have been designed to self-check your progress of study. It will be better if you solve the problems put in these boxes immediately after you go through the sections of the units and then match your answers with 'ANSWERS TO CHECK YOUR PROGRESS' given at the end of each unit.In the previous block of this course, we have already studied seven units out of a total of fifteen units. In this block, we shall study the remaining eight units.

UNIT 1: BASIC CONCEPT IN SAMPLING

UNIT STRUCTURE

1.1 Learning Objectives

1.2 Introduction

1.3 Defining Population, Parameter and Sampling

- 1.3.1 Random and Non-Random Sampling Methods
- 1.3.2 Sampling with and without Replacement
- 1.3.3 Large Samples

1.4 Errors in Sampling Procedure

1.5 Let Us Sum Up

1.6 Further Reading

1.7 Answer to Check Your Progress

1.8 Model Questions

1.1 LEARNING OBJECTIVES

After going through this unit, you will be able to:

- discuss the concept of a parameter and a statistic
- explain the concepts of a sampling with and without replacement
- explain random and non-random sampling
- discuss about standard error
- discuss about different methods of sampling
- explain merits and limitations of sampling.

1.2 INTRODUCTION

Often we are interested in drawing valid conclusions about a large group of individuals or objects. Instead of examining the entire group, which may be difficult to do, we examine only a small part. We do this with the aim of inferring certain facts about the group of individuals.

The entire group of individuals is known as population. The word ‘population’ used in statistics has altogether a different meaning.

You all know the dictionary meaning of the word population. But in

mathematics we use the word with a bit different meaning. Mathematician use the word population to refer not only to people but to all items that have been chosen for Study. When it is possible to examine every item in the population that we have taken for study, then it is called a Census. But complete enumeration is not always possible in all situations. Then we take another procedure, which we will refer as sampling method. Mathematician use the word sample to describe portion chosen from the population. For example we have population of heights of all female inhabitants of Guwahati or a population of incomes of families living in Guwahati etc. A population may be finite, infinite or hypothetical (for example, outcomes of tossing a coin infinite number of times).

Often you have seen in news paper that Government estimated the production of rice in a year. Do they take complete enumeration? Certainly not. When they do this they select some areas and measure the production there and after that they do some mathematical calculation. This is just an example. Every estimation is done like this. Care should be taken in selecting the areas; otherwise you will not get good result. Sample must be the representative of population.

1.3 DEFINING POPULATION, PARAMETER AND SAMPLING

In mathematics, we describe samples and population by using measures such as mean, median, mode and standard deviation (s.d). When these terms describe the characteristics of a sample then they are called ***Statistics***. A 'Statistic' is a function of the sample. Examples are arithmetic mean $\left(\bar{x}\right)$ and sample variances (σ^2).

When they describe the characteristics of a population they are called ***Parameters***.

A ***population*** is the totality of any kind of units under consideration. It may be finite or infinite.

A sample is any portion of the population selected for study. The process of obtaining samples is called ***sampling***.

The fundamental assumption underlying the theory of sampling is random sampling, which consists in selecting the individuals from the populations, in such a way that each individuals of the population has the equal chance of being selected.

Two types of sampling will be discussed– (1) Random Sampling and (2) Non random Sampling.

Purpose of Sampling: A sample is not studied for its own sake. The main objective of its study is to draw inference about the population. In other words sampling is only a tool which helps us to know the characteristic of the universe or population by examining only a small part of it.

Simple Sampling: Sampling of attributes may be regarded as the drawing of samples from a population whose members possess the attribute A or not (–A). The choosing of an individual in a sampling is called an event or trial. Possession of a specified attribute A by the individual is called a success.

Simple sampling means random sampling in which each event has the same probability ‘‘p’’ of success and the probability of success is independent of the success or failure of the events in the preceding trials. Thus simple sampling is a special case of random sampling in which trials are independent and probability of success is constant.

Let us consider an example. In an urn there are 2 white and 3 black balls. Probability of drawing a white ball is $\frac{2}{5}$. If a white ball is drawn on the first trial, probability of drawing a white ball will be $\frac{2}{5}$. In the second trial if the ball is not replaced the probability of drawing a white ball will be $\frac{1}{4}$. Hence the sampling though random is not simple.

A random sampling from an infinite population is always a simple since drawing of one individual cannot affect the probability. But a sampling from finite population may or may not be simple.

Let us take a simple sample of ‘n’ members, which will be identical with a series of n independent trials with probability of success p (constant).

The probabilities of 0, 1, 2, …, …, …, n success are the terms of binomial expansion of $(p+q)^n$ which can be written in sequence. The distribution so obtained is called sampling distribution of number successes in the sample. The mean value and S.d (standard error S.E.) are (Missing) np and $\sqrt{npq}$ respectively.

Now proportion of successes in a sample is the number successes divided by the number of members in the sample. Hence,

Mean of proportion of success = p

(S.E) of proportion of success = $\sqrt{\frac{pq}{n}}$

Already we have mentioned that care should be taken in selecting samples, otherwise inference made from the result get will mislead.

There are two types of sampling:

i) Random sampling

ii) Nonrandom sampling.

In random sampling all the items in the population have equal chance of being chosen in the sampling. Whereas in nonrandom sampling personal knowledge and opinion are used to identify the items from the population that are to be included in the Sample. A Sample selected by this sampling is based on someone's expertise. The rigorous statistical analysis that can be done by random sample but cannot be done by non-random sample. Look to the following table for clear understanding the distinction between populations and samples.

	Population	Sample
Definition	Complete enumeration of items is considered	Part of the population chosen for study
Characteristics	Parameters	Statistics
Symbols	Population size = N Population mean = μ Population S.d = σ	Sample size = n Sample mean = $\bar{x}$ Sample S.d = s

In random sample, the probability of an element of the population will or will not be included in the Sample are known. As a result we can

assess mathematically the estimates of the population characteristics from the result of the Sample.

1.3.1 Random and Non-Random Sampling Methods

Sampling is a process of studying the characteristics of a population by selecting samples from it. Let us discuss some random and non-random methods–

Methods of Random Sampling: When a sample is drawn from a population, it is necessary to decide which method should be applicable. The various methods of sampling can be grouped under the heads as **random sampling** and **non-random sampling.**

Some of the sampling methods are as follows:

A) Random Sampling Method:
   i) Simple Random Sampling
   ii) Systematic Sampling
   iii) Stratified Sampling
   iv) Cluster Sampling

B) Non-random Sampling Methods:
   i) Judgment Method
   ii) Quota Sampling
   iii) Convenience Sampling

A) Random Sampling Method:

i) Simple Random Sampling: This method selects samples that allows:

a) Each possible sample to have an equal probability of being selected.

b) Each item in the population has equal probability of being included in the Sample.

Let n units of the sample be drawn from the population one by one, and after drawing the item selected is returned to the population before the next drawing is made. In this case at each stage of drawing, each of N items of the population gets the same probability $\frac{1}{N}$ of being selected.

This sampling procedure is called Simple random Sampling with replacement. In this case, the same unit of the population may occur more than once in the Sample. There will be n possible Samples and each has the probability $\frac{1}{N^n}$.

In other case, the n members of the Sample are drawn one by one, but at any stage members drawn in the previous drawing is not returned to the population. This procedure is known as Simple random Sampling without replacement.

In this case no member of the population can occur more than once in the sample. There are Ncn possible Samples and each Sample has the probability.

$$\frac{n}{N}.\frac{n-1}{N-1}.\frac{n-2}{N-2}.\frac{n-3}{N-3}.---.\frac{1}{N-n+1}=\frac{1}{N_{C_n}}$$

In this case also the probability that any particular population units will be selected of the drawing is:

$$\frac{n-1}{N}.\frac{n-2}{N-1}.\frac{n-3}{N-2}.---.\frac{N-r+1}{N-r+1}.\frac{1}{N-r+1}=\frac{1}{N}$$

Technique of Simple Random Sampling: The easiest way to select a Sample randomly is to use random numbers. These numbers can be taken from a table of random digits or can be generated by a computer programming.

For example, suppose a company has 1000 employees and wish to interview a randomly chosen sample of 10. Write the name of each one on a slip of paper and put the papers on a cylindrical drum and rotate it so that they are thoroughly mixed. Now draw 10 slips at random.

ii) **Systematic Sampling:** In these sampling elements are selected from the population at a uniform interval that is measured in time, space, order or any others.Here each element in the population has the same chance of being selected from the sample. For example suppose we want to interview 20th students on a college campus. We choose a

random starting point in the first 20 names in the student register book and then pick every 20th name thereafter. In this sampling also each element has an equal chance of being selected; but each sample does not have an equal chance of being selected. In this point this sampling differs from Simple random Sampling. In this sampling there is problem of introducing error into the Sample process. The only advantage of this sampling is that it takes less time and lower cost.

iii) **Stratified Sampling:** In this Sampling, population is divided into homogeneous groups. These groups are called strata. There are two approaches:

a) From each strata, select specified number of elements at random.

b) Draw equal number of elements from each strata and give weight to the result according to the Stratum's proportion in the population. In Stratified sampling, every element in the population has chance being selected. In this sampling each group has small variation within itself. There are wide variations between the groups.

iv) **Cluster Sampling:** In this sampling also population is divided into groups and the groups are called Clusters. The assumption is that individual clusters are representative of population. Then random Samples are selected from the clusters. In case of cluster sample there is considerable variation within each group, but the groups are essentially similar to each other. In cluster, stratified and systematic samples attempts are made to approximate simple random sample. These three methods are developed for economical and physical considerations. Simple random sampling is basis of all the samplings and it is fundamental one. Principles of simple random sampling are the foundations of statistical inferences.

B) Non-Random Sampling Methods:

i) Judgment Method: In this method the choice of sample items depends exclusively on the judgment of the investigator. For example, if a sample of 10 artist is to be selected from a group of 50 artists, then the investigator would select 10 artists who, in his opinion, represent the group.

Though this method is simple, it is not scientific because the result may be considerably affected by the personal biasness of the investigator. This method is generally applicable when the size of the sample is small. In general, for solving business problems, making public policy decisions, pilot studies etc we use this method.

ii) Quota Sampling: It is a type of judgment sampling where quotas are set up according to given criteria but within the quotas the selection of sample items depends on personal judgment. For example, if you want to take interview of 100 people, out of which there should be 50 Hindu, 50 Muslim, 50 Sikhs and 50 Christian. Within these quotas, the interviewer is free to select the people interviewed.

iii) Convenience Sampling: In this method the sample is obtained from readily available lists, such as telephone directories, or automobile registration are example of convenience sampling.

Merits of Sampling Method:

1) It is less time consuming compared to census.
2) It is less costly either in terms of money or in terms of man powers, because only a small portion of the total population is studied.
3) More detailed information can be gathered as the size of the sample is small compared to total population.
4) A sample survey generally gives data of better quality than a complete census, because in a sample survey it

may be possible to employ better trained personnel, effect better supervision or use better equipment than is possible or feasible in a complete census.

5) For infinite population, sampling is the only solution.

Limitations of Sampling:

1) Without proper planning, the samples result may be inaccurate and misleading.
2) In absence of qualified persons, the information obtained from samples surveys can not be relied upon.
3) If the information is required from whole population then sampling is meaningless.



CHECK YOUR PROGRESS

Q.1: Fill in the blanks with appropriate words:

i) In ..... sampling all the items in the population have equal chance of being chosen in the sampling.

ii) In ..... sampling personal knowledge and opinion are used to identify the items .

iii) Convenience sampling is a method of ..... sampling.

iv) In ..... Sampling population is divided into homogeneous groups.

v) Sampling is less time consuming compared to .....

1.3.2 Sampling with and without Replacement

Sampling where each member of the population may be chosen more than once is called a sample with replacement, while sampling where each member cannot be chosen more than once is called as sample without replacement.

If we draw an object from an urn, we have the choice of replacing or not replacing the object into the urn before we draw

again. In the first case a particular object can come up again and again, whereas in the second case it can come up only once.

Random Sampling: If in a sampling each member of the population has the same chance of being selected then it is called a random sampling.

A random sample of size 'n' taken with replacement from a population with distribution p(x) = P (X = x) is a collection of random variables x_1, x_2, …, …, ..., x_n each having the same distribution as x and x_1, x_2, …, ..., ..., x_n are independent.

Then μ = Mean of the population = $\frac{\sum x_i p_i}{n}$ = E(n)

Let us denote N numbers of the elements of the set be x_1, x_2, ..., ..., ..., x_n, taking a sample consisting of n elements from the set of N numbers. Sample vatiables are x_1, x_2, ..., ..., ..., x_n.

Mean of population $\mu = \frac{\sum_{i=1}^{N} x_i}{N}$

Mean of sample $\bar{x} = \frac{\sum x_i}{n}$

Now $E(\bar{x}) = E\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(x_i)$ ... ... (1)

$$E(x_i) = \frac{X_1}{N} + \frac{X_2}{N} + ... + \frac{X_n}{N}$$

[As x_i can take any element from X_1, X_2, ..., ..., ..., X_N and its probability is $\frac{1}{N}$]

$\therefore E(x_i) = \frac{1}{N}\sum_{i=1}^{N} x_i = \mu$ ... ... (2)

From (1) and (2): $E(\bar{x}) = \frac{1}{N}\sum_{i=1}^{n} \mu = \frac{1}{N}.n\mu = \mu$

Variance $\sigma = \frac{1}{N}\sum_{i=1}^{N} (X_i - \mu)^2$

If all probabilities are equal and equal to $\frac{1}{N}$ then,

Mean = $\frac{\sum x_i}{n}$ = E(x)

Variance = Var(x) = σ^2 = $E(x^2) - [E(x^2)]^2$

= σ^2

Where $[E(x)]^2 = \sum_{i=1}^{n} x_i^2 p(x_i)$

EXERCISE 1.1

A population consists of five numbers 2, 3, 6, 8 and 11. Consider all possible samples of size 2 that can be drawn with replacement from this population. Find mean and s.d. of the population. Also find mean and s.d. of the samples.

Solution: Let μ = Mean of population

$$= \frac{2+3+6+8+11}{5}$$

$$= \frac{30}{5}$$

$$= 6$$

Again $\sigma^2 = \frac{(2-6)^2+(3-6)^2+(6-6)^2+(8-6)^2+(11-6)^2}{5}$

$$= \frac{16+9+0+4+25}{5} = 10.8$$

Therefore Standard Deviation of the population = σ

$$= \sqrt{10.8}$$

$$= 3.29$$

The samples of size of that can be drawn with replace mean are:

(2, 2)	(2, 3)	(2, 6)	(2, 8)	(2, 11)
(3, 2)	(3, 3)	(3, 6)	(3, 8)	(3, 11)
(6, 2)	(6, 3)	(6, 6)	(6, 8)	(6, 11)
(8, 2)	(8, 3)	(8, 6)	(8, 8)	(8, 11)
(11, 2)	(11, 3)	(11, 6)	(11, 8)	(11, 11)

They are 25 in numbers the corresponding sample means are

2.0	2.5	4.0	5.0	6.5
2.5	3	4.5	5.5	7.0
4.0	4.5	6.0	7.0	8.5
5.0	5.5	7.0	8.0	9.5
6.5	7.0	8.5	9.5	11.0

Mean of the sampling distribution is:

$$\mu_{\bar{x}} = \frac{\text{Sum of the sample means}}{25}$$

$$= \frac{150}{25} = 6.0$$

So $\mu_{\bar{x}} = \mu$

For S.D of samples, subtract 6 from all the means of the sample and square them.

Add all squared members and divide the result by 25.

$$\sigma^2_{\bar{x}} = \frac{135}{25} = 5.4$$

S.D. of samples = $\sigma_{\bar{x}}$

$$= \sqrt{5.4}$$

$$= 3\sqrt{.6} = 2.32$$

Let us consider the same problem without replacement.

Here also, $\mu = 6$

And $\sigma = 3.29$

The samples in this case will be (2,3),(2,6), (2,8), (2,11), (3,6), (3,8),(3,11), (6,8), (6,11) and (8,11).

They are 10 in numbers.

The corresponding sample means are 2.5, 4, 5, 6.5, 4.5, 5.5, 7, 7, 8.5 and 9.5.

Mean of the sampling means are:

$$\mu_{\bar{x}} = \frac{2.4 + 4 + 5 + 6.5 + 4.5 + 5.5 + 7 + 7 + 8.5 + 9.5}{10}$$

$$= \frac{60}{10} = 6$$

So, $\mu_{\bar{x}} = \mu$

Variance of the sample mean is:

$$\sigma^2_{\bar{x}} = \frac{(2.5-6)^2+(4-6)^2+(5-6)^2+(6.5-6)^2+(4.5-6)^2+(5.5-6)^2+(7-6)^2+(8.5-6)^2+(9.5-6)^2}{10}$$

$$= \frac{40.5}{10} = 4.05$$

$$\sigma_{\bar{x}} = 2.01$$

1.3.3 Large Samples

Suppose that a large number n of independents Bernoull's trials is performed and x successes are obtained. The hypothesis we want to test is that the probability of success in each trial is p. Assuming the hypothesis to be correct, the mean and variance of the sampling distribution of the number of success are np and npq.

Again we know for large n, $Z = \frac{x - np}{\sqrt{npq}}$ is distributed as a standard normal variate.

From table $P(|Z| > 3)$ = 0.0027. Therefore we conclude that the hypothesis is improbable if the difference between the observed number of successes x and the expected number of successes np, is highly significant.

If $|Z|$ < 1.96 the data is consistent with the hypothesis.

If $|Z|$ > 1.96 the difference is significant.

Let the observed proportion of success in a sample of n observation be p then,

Mean = p and variance = $\frac{pq}{n}$

Thus $\frac{\frac{x}{n-p}}{\sqrt{\frac{pq}{n}}}$ ~ N(0, 1)

This test is valid for large n only.

Standard error: The s.d. of the Sampling distribution of statistic is known as Standard error (SE). If n be the sample size and σ^2 is the population variance then SE of the statistic $\bar{x}$ of the sample is $\frac{\sigma}{\sqrt{n}}$.

SE plays a very important role in the large sample theory and forms the basis of the testing of hypothesis.

If t be any statistic then for large samples,

$$Z = \frac{t - E(t)}{S.E(t)} \sim N(0, 1)$$

Thus if the discrepancy between the observed and expected value of the statistic Z is greater than Z. S.E, then hypothesis is rejected at certain level of significance.

1.4 ERRORS IN SAMPLING PROCEDURE

There are two types of errors in sampling procedure. These are as follows:

1) Sampling error and
2) Non-sampling error

1) Sampling error: A sample covers only a portion of the population from which it is drawn. The information gathered through sample is used for drawing certain inferences about the characteristic of the population. A discrepancy may arise between the population parameter and the sample statistic. The error thus occurs is called sampling error. There are two factors which will lead to sampling error. They are sampling bias and chance factor. When the method of selection is defective, sampling bias will be caused. This error can be minimized by use of proper sampling method. Even when sample bias is absent, discrepancy may occur due to chance.

2) Non-sampling error:

- **Response error:** It is due to response bias- viz vague answer, wrong answer and inaccurate answer given by a respondent.
- **Observational error:** It is due to bias observation.

- **Errors due to non response:** An interviewer may not be able to get information from a respondent.

Test of Significance: A very important aspect of Sampling theory is the study of the test of significance, which enables us on the basis of the Sample results, if the deviation between the observed sample statistic and the hypothetical parameter value.

The technique of randomization used for the selection of Sample units makes the test of significance valid. For applying the test of significance we first set up the hypothesis– a definite statement about the population parameter. This hypothesis is called null hypothesis.



EXERCISE 1.2

A dice is thrown 9,000 times and a throw of 3 or 4 observed 3,240 times. Show that the dice is biased.

Solution: Here, N = 9,000

X = number of successes = 3,240

Hypothesis: Dice is unbiased

P = probability of getting 3 or 4 = $\frac{1}{6}+\frac{1}{6}$

$= \frac{1}{3}$

Expected mean = np = $9{,}000 \times \frac{1}{3}$

= 3,000

n is large $\Rightarrow |Z| = \frac{x-np}{\sqrt{npq}}$

Standard Error (S.E.) = $\sqrt{npq}$

$= \sqrt{9000 \times \frac{1}{3} \times \frac{2}{3}}$

$= \sqrt{2000}$

$= 20\sqrt{5}$

$$Z = \frac{3240 - 3000}{20\sqrt{5}}$$

$$= \frac{240}{20\sqrt{5}}$$

$$= \frac{12\sqrt{5}}{5}$$

$$= 24\sqrt{5}$$

$$= 5.36$$

Since $|Z| > 3$, the dice is biased.



EXERCISE 1.3

A random sample of 500 pineapples was taken from a large contingent and 65 were found to be bad. Show that Standard Error of the population of bad ones in a sample of this size is 0.015.

Solution: Here, n = 500

x = number of bad pineapples in the sample = 65

$$p = \frac{65}{500} = 0.13$$

$$q = 1 - 0.13$$

$$= 0.87$$

$$\text{S.E.} = \sqrt{\frac{pq}{n}} = \sqrt{\frac{0.13 \times 0.87}{500}} = 0.015$$



CHECK YOUR PROGRESS

Q.2: In a sample of 1000 people in Guwahati, 840 are rice eaters and rest is wheat eaters. Calculate Z.

.....



1.5 LET US SUM UP

In this unit we have developed the concept of random sampling and discuss simple sample as a particular case of random sample. We have also discussed sampling with and without replacement. In all these discussion we have taken sample from the normal population.

We have also discussed about different methods of sampling - simple random sampling, systematic sampling, stratified sampling and cluster sampling (all these are random sampling). Again we have discussed about some non random sampling like- judgment method, quota method and convenience sampling. We have also discussed about sampling and non sampling errors.



1.6 FURTHER READING

1) Agarwal, D. R. (2006). *Business Statistics*. Delhi: Vrinda Publications.

2) Gupta, S. C. (1994). *Fundamentals of Statistics*. New Delhi: Himalayan Publishing House.

3) Rajagopalan, S. P. & Sattanthan, R. (2009). *Business Statistics and Operations Research*. New Delhi: Tata McGraw-Hill.

4) Sharma, J. K. (2007). *Business Statistics*. New Delhi: Pearson Education Ltd.

5) Verma, A. P. (2007). *Business Statistics*. Guwahat: Asian Books Private Limited.



1.7 ANSWERS TO CHECK YOUR PROGRESS

Ans. to Q. No. 1: i) random sampling, ii) non random sampling, iii) non-random sampling, iv) Stratified sampling, v) census.

Ans. to Q. No. 2: 24.64.



1.8 MODEL QUESTIONS

Q.1: A population is made up of groups that have wide variation within each group but little variation from group to group. What will be the appropriate type of sampling for this population?

Q.2: List the advantages of sampling over census.

Q.3: A population consists of four members 3, 7, 11 and 15. Find all possible samples of size 2 that can be drawn with replacement from this population. Find population mean and population s.d. Find also mean of sampling mean and s.d. of sampling s.d.

Q.4: Solve the above problem without replacement.

Q.5: The masses of 1500 ball bearings are normally distributed with mean 22.40 kg. and s.d. of 0.048 kg. If 3000 random samples of size 36 are drawn from this population, determine the mean and s.d. of the sampling means if sampling is done (a) with replacement (b0 without replacement.

Q.6: In a random sample of 500 people from a particular district, 300 are found to be smokers. In one of 1000 people from another district, 550 are smokers. Calculate Z.

Q.7: A coin is tossed 10,000 times and it turns up head 5,195 times. Can the coin be regarded as unbiased?

Q.8: In some dice throwing experiments, 49,152 times of 75,145 yielded 4, 5, or 6. Is the dice unbiased?

Q.9: A random sample of 500 apples was taken from a large consignment and 60 were found to be bad. Calculate Standard Error.

Q.10: A sample of 900 members has mean 3.4 cm and s.d. 2.61 cm Is the sample from a large population of mean 3.25 cm and s.d. 2.61 cm?

*** ***** ***

UNIT 2: SAMPLING TECHNIQUES - I

UNIT STRUCTURE

2.1 Learning Objectives
2.2 Introduction
2.3 Random Sampling
2.4. Stratified Random Sampling
2.5 Systematic Sampling
2.6 Cluster Sampling
2.7 Let Us Sum Up
2.8 Further Reading
2.9 Answer to Check Your Progress
2.10 Model Questions

2.1 LEARNING OBJECTIVES

After going through this unit, you will be able to-

- put forward the meaning of sampling techniques
- describe the different types of sampling techniques
- describe the examples of different types of sampling techniques.

2.2 INTRODUCTION

Basically, there are two types of sampling methods, viz., probability sampling and non-probability sampling. In this unit, we shall discuss probability sampling and its different types.

In the probability sampling, selection of the each and every unit of the population into sample is determined only by the probability or chance, and does not depend on any other considerations. In non-probability sampling selection of units of sample does not depend on probability or change but on other important considerations such as specific purpose and requirement of the study, judgement of the researchers, convenience of selecting the sample etc. The most importantly, probability sampling allows reasonable generalisations about the population parameters. Contrarily, generalisations are not carried out on the basis of non-probability samples.

2.3 SIMPLE RANDOM SAMPLING

Among the probability sampling methods, Simple Random Sampling (SRS) is the most popular and common one. In Simple Random Sampling every sampling unit of the population has an equal chance of being selected to the sample. Equal probability here implies that the each element in the population has got the same chance of being selected to the sample. The Simple Random Sampling preferably requires the following for effective implementation:

- Complete list of units to be studied should be available;
- Size of the universe should be known;
- Sample size should be adequately specified;
- Each unit should be ensured equal chance of being selected to the sample.

  Let us now discuss how to draw a simple random sample along with its different types. Simple Random Sample is drawn by the following methods:

- **Lottery method:** Here, firstly the researcher needs to prepare the complete list of all the units of the population. Secondly, all the units are to be written down in small pieces of paper. Next, all the pieces of paper to be put in a jar and mix well. Finally, the researcher has to select papers from jar the one by one and continue till the desired numbers of samples are not fulfilled. For example, suppose, a researcher wants to study the relationship between income and expenditure of the people in a certain village consisting of 300 households, with house numbers allotted from 1 to 300. The researcher then first has to write the house numbers on pieces of papers, then put them into a jar and accordingly select the slips with house numbers written on it. If the researcher wants to have a sample size of 100 households, he or she needs to pick 100 slips of papers and the numbers written on it will be the respective units to be selected.
- **Tippet table or table of random numbers:** The statisticians have prepared a set of tables called Tippet Table or Tables of Random

numbers. A sample is framed with reference to these tables. Of these tables, Tippet's Table is most widely used. Using 41,600 figures, Tippet has involved 10,400 numbers comprising of four units each. In using this method, all items of the universe are first arranged in an order and then by using Tippet's Table the required numbers of items are selected as per the requirement of the sample size.

Sampling with or without replacement:

Random sampling can be selected using two different systems: sampling without replacement and sampling with replacement. Suppose we want to select a sample of 20 students out of a total of 80. The first student is selected out of the total class, and so the probability of selection for the first student is 1/80. When you select the second student there are only 79 left in the class and the probability of selection for the second student is not 1/80 but 1/79. The probability of selecting the next student is 1/78. By the time you select the 20th student, the probability of his/her selection is 1/61. This type of sampling is called sampling without replacement. But this is contrary to our basic definition of randomisation; that is, each element has an equal and independent chance of selection. In the second system, called sampling with replacement, the selected element is replaced in the sampling population and if it is selected again, it is discarded and the next one is selected. If the sampling population is fairly large, the probability of selecting the same element twice is fairly remote. In most of the cases the second one i.e. sampling with replacement is adopted.

Advantages of Simple Random Sampling: The advantages of simple random sampling include the following:

- All elements of the population have equal chance of being selected in the sample;
- This method is free from personal bias of the investigator;
- It is the simplest of all the sampling methods and easier to conduct;
- This method can be used in conjunction with other methods in probability sampling;
- Degree of sampling error is low;
- The universe gets fairly represented by the sample drawn using this

method.

Disadvantages of Simple Random Sampling: However, there are some disadvantages of simple random sampling like:

- It does not make use of the knowledge of population which the researcher may have;
- Sometimes, operationalisation of simple random sampling method, especially when an exhaustive list of the population units is absent, becomes very difficult.
- It cannot be used if the researcher wants to break the respondents into subgroups for comparison purposes.

2.4 STRATIFIED SAMPLING

This method of sampling is generally used when the population does not constitute a homogeneous group. According to this method of sampling, the population is divided into several sub-population or sub-groups called 'strata' having common characteristics. From each stratum, items are selected to make the sample. Since each stratum is more homogeneous than the total population, we are able to get more precise estimates for each stratum and by estimating more accurately each of the component parts; we get a better estimate of the whole. In brief, stratified produces better representations on several occasions. It is also important for the characteristic that becomes the basis of stratification to be related to the main variable that you are exploring. Suppose we want to conduct a research upon the income expenditure behaviour among the employees of an office. Here, we must look for these three relevant questions in the context of stratified sampling:

- How to form the strata?
- How should the items be selected from each stratum?
- How many items are to be selected from each stratum or how to allocate the sample size of each stratum?

Let us now briefly discuss these questions.

- *How to form the strata?* To resolve this question, the researcher can form the strata on the basis of the common characteristics. For this,

the entire population should be divided into sub groups with homogeneous nature. For the example cited above, the employees have different income groups based on certain condition. So, we can arrange the employees according to their income ranges say between Rs.10,000-20,000, 20,000-30,000, 30,000-50,000, etc. In this manner, we will divide the entire population of employees of the office into the smaller groups with a common characteristic being their monthly income.

- *How should the items be selected from each stratum?* For the second question, the most usually used method of selecting units from the stratum for the sample is simple random sampling. There are two types of stratified sampling: proportionate stratified sampling and disproportionate stratified sampling. With proportionate stratified sampling, the number of elements from each stratum is selected as per the proportions to the total population, whereas in disproportionate stratified sampling, consideration is not given to the size of the stratum.
- *How many items are to be selected from each stratum?* Regarding the third question cited above, generally proportionate allocation is followed. If Pi represents the proportion of population included in stratum i, and n represents the total sample size, the number of elements selected from stratum i is n.Pi. Suppose we want a sample of size n=20 to be drawn from a population of size N=100 which is divided into three strata with 30, 50 and 20 elements in each. Using the proportionate method, number of units to be selected from the three strata would be 6, 10 and 4 units respectively.

Advantages of Stratified Sampling:

- The sample chosen can represent various groups and patters of characteristics in the desired proportions.
- It can be used for comparing sub-categories.
- It provides reliable and meaningful results.

Disadvantages of Stratified Sampling:

- It requires more time and efforts than simple random sampling.
- It is a comparatively complex method of selecting sample.

- This method is suitable only when complete knowledge about the diverse characteristics of the population is available. Hence, it has a limited scope.
- There is possibility of biasness while classifying the population into different strata.
- If size of population is already small, it may be difficult to further divide it into smaller groups or strata.

2.5 SYSTEMATIC OR INTERVAL SAMPLING

According to this method, units of the population are numerically, geographically or alphabetically arranged and then every nth item of the population is selected for the sample. In simple terms, it is randomly selecting the first respondent and then every nth person after that; where 'n' is called the sampling interval.

For instance, if we want to collect a sample of 10 individuals out of 100, then first we would systematically arrange the people. Secondly, any one individual from the first 10 individual would be randomly selected. After that every tenth person would be selected for the sample. Like, if we first select the 5th person, then the sample would include the 5th, 15th, 25th, 35th, 45th, 55th, 65th, 75th, 85th, 95th individual. Thus, in systematic sampling only the first unit is selected randomly and the remaining units of the sample are selected at fixed intervals. The sampling interval is taken as N/n, where N is the population size from which 'n' number of units are to be selected as sample. Systematic sampling differs from simple random sampling as in the latter selection of sample units are independent of each other, whereas in the former the selection of sample units depend upon the selection of a previous unit.

Advantages of Systematic Sampling:

- It is considered as an improvement from a simple random sample as a systematic sample is spread more evenly over the entire population.
- It is an easier and less costly method of sampling.
- It can also be used when size of population is large. There is hardly any possibility of personal bias in this method.

- It is a rapid method and eliminates several steps otherwise taken in the other methods.

Disadvantages of Systematic Sampling:

- Every item in the population does not get the equal chance of being selected because only the first item is selected randomly.
- It ignores the units of population between two nth numbers. This may result in possibility of over representation or under representation of several groups.
- If there is some hidden periodicity in the population, systematic sampling will prove to be inefficient method of sampling. In practise, systematic sampling is efficient when list of population is available and they are of considerable length.



CHECK YOUR PROGRESS

Q 1: Mention the other name of probability sampling. Answer in about 10 words.

.....

.....

.....

.....

.....

Q 2: What is meant by stratified sampling? Answer in about 40 words.

.....

.....

.....

.....

Q 3: What is meant by systematic sampling? Answer in about 40 words.

.....

.....

.....

.....

2.6 CLUSTER SAMPLING

Simple random and stratified sampling techniques are based on a researcher's ability to identify each element in a population. It is easy to do this if the total sampling population is small, but if the population is large, as in the case of a city, state or country, it becomes difficult and expensive to identify each sampling unit. In such cases the use of cluster sampling is more appropriate.

Cluster sampling is based on the ability of the researcher to divide the sampling population into groups (based upon visible or easily identifiable characteristics), called clusters, and then to select elements within each cluster, using the simple random sampling technique. Clusters can be formed on the basis of geographical proximity (called area sampling) or a common characteristic that has a correlation with the main variable of the study (as in stratified sampling). Depending on the level of clustering, sometimes sampling may be done at different levels. These levels constitute the different stages (single, double or multiple) of clustering.

Suppose, a researcher wants to study the problems faced by the undergraduate students in India in the context of pursuing their higher education. UG students are to be found in ample all over the country. Moreover, there are numerous institutions both colleges and universities that offer UG courses. In every institute there are various disciplines that provide UG courses that generally cover a period of three to four years. Thus, the magnitude of the task on the part of the researcher can be well understood. In such situations, cluster sampling is extremely useful in selecting a random sample.

The first level of cluster sampling could be at the state or territory level. Clusters could be grouped according to similar characteristics that ensure their comparability in terms of student population. If this is not easy, you may decide to select all the states and territories and then select a sample at the institutional level. For example, with a simple random technique, one institution from each category within each state could be selected (one university and one college). This is based upon the assumption

that institutions within a category are fairly similar with regards to student profile. Then, within an institution on a random basis, one or more academic programmes could be selected, depending on resources. Within each study programme selected, students studying in a particular year could then be selected. Further, selection of a proportion of students studying in a particular year could then be made using the simple random technique. The process of selecting a sample in this manner is called multi-stage cluster sampling.

Advantages of Cluster Sampling:

- It is convenient to use cluster sampling when population is large or a large geographical area is taken for the study.
- The cost involved in this method is relatively less than the other methods of sampling.
- Respondents can be readily substituted for other respondents.
- It ensures flexibility in sampling.
- Characteristics of clusters can be estimated.
- It can be used when it is inconvenient to randomly select respondents.

Disadvantages of Cluster Sampling:

- Each cluster is not of equal size. For example while having clusters of towns and villages, each town or each village in the country is not of equal size.
- Comparison would not be on equal basis.
- Sampling error would be greater.
- In some situations, one individual might belong to two clusters and studied twice.
- It lacks representation.
- There could be homogeneity in one cluster but heterogeneity in the other.

Difference between Cluster Sampling and Stratified Sampling:

The difference between cluster sampling and stratified sampling is that in the former homogeneous group (say a geographical region) is classified into heterogeneous units (towns, cities, villages, etc.) whereas in the latter a heterogeneous group is classified into homogeneous sub groups called strata.

Area Sampling: When clusters happen to be based on some geographical subdivision, it is called area sampling. For example, suppose you want to study the impact of government schemes on the entrepreneurs of Kamrup district. For this you have to divide the entire district into sub groups like blocks and then from each block select random samples depending upon the sample size.

Multi-stage Sampling: Multi-stage sampling is a further development of the principle of cluster sampling. In this method, sampling is selected in various stages but only the last sample of subjects is studied. For example, if we want to study the status of self-help-groups in India, first we will divide the entire country into four zones East, west, north and south. Then from each zone, we will select two states randomly; say from east we have Assam and Kolkata, from West we have Gujarat and Maharashtra, from north we have Punjab and Himachal Pradesh and from the south we have Kerala and Tamil Nadu. Next from each of these 8 states we will randomly select two towns and two villages randomly. Out of these 16 selected towns and villages we will randomly select a sample of say 1000 SHGs. Now, these 1000 respondents would be the representative of the entire country and with their data we will conduct the study about the status of SHGs in the nation.

Thus, multi-stage sampling is generally applied in big enquiries extending to a considerably large geographical area, say the entire country. Since sampling frame under multi-stage sampling is developed in partial units and a large number of units can be sampled for a given cost, this method of sampling is considered to be advantageous than the other ways of sampling.



CHECK YOUR PROGRESS

Q 4: What is meant by cluster sampling. Answer in about 50 words.

.....

.....

.....

.....



2.7 LET US SUM UP

- Basically, there are two types of sampling methods- probability sampling and non-probability sampling. In this unit, you will learn about the former and its types. In probability sampling every unit of the population has an equal and independent chance of being selected for the sample. Whereas, in non-probability sampling every unit does not have the equal chance of getting selected in the sample, rather the selection of units for the sample depends upon the researcher.
- If sample size selected for study is very small or due to non-representative nature of the sample, the estimated results may differ from the actual results. The error thus emerging is called sampling error. Sampling errors can be reduced by increasing the sample size. Simple random sampling, stratified sampling, systematic sampling and cluster sampling are the types of probability sampling.
- The stratified method of sampling is generally used when the population does not constitute a homogeneous group. According to this method of sampling, the population is divided into several sub-population or sub-groups called 'strata' having common characteristics. From each stratum, items are selected to make the sample for the study.
- According to the Systematic or interval sampling method, units of the population are numerically, geographically or alphabetically arranged and then every nth item of the population is selected for the sample. In simple terms, it is randomly selecting the first respondent and then every nth person after that; 'n' is a number termed as sampling interval. If the population is large, as in the case of a city, state or country, it becomes difficult and expensive to identify each sampling unit. In such cases the use of cluster sampling is more appropriate. Cluster sampling is based on the ability of the researcher to divide the sampling population into groups called clusters, and then to select elements

within each cluster, using the simple random sampling technique. Clusters can be formed on the basis of geographical proximity (called area sampling) or a common characteristic that has a correlation with the main variable of the study (as in stratified sampling. Depending on the level of clustering, sometimes sampling may be done at different levels.

- Multi-stage sampling is a further development of the principle of cluster sampling. In this method, sampling is selected in various stages but only the last sample of subjects is studied.



2.8 FURTHER READING

1) Ahuja, R. (2018). Research Methods. Jaipur: Rawat Publications.

2) Kothari, C.R. (2011). Research Methodology, Methods and Techniques. New Age International (P) Ltd.

3) Kumar, R. (2011). Research Methodology. New Delhi: SAGE Publications India Pvt. Ltd.



2.9 ANSWERS TO CHECK YOUR PROGRESS

Ans to Q No 1: Probability sampling is also known as 'random sampling' or 'chance sampling'.

Ans to Q No 2: The stratified method of sampling is generally used when the population does not constitute a homogeneous group. According to this method of sampling, the population is divided into several sub-population or sub-groups called 'strata' having common characteristics. From each stratum, items are selected to make the sample for the study.

Ans to Q No 3: According to the Systematic or interval sampling method, units of the population are numerically, geographically or

alphabetically arranged and then every nth item of the population is selected for the sample. In simple terms, it is randomly selecting the first respondent and then every nth person after that; 'n' is a number termed as sampling interval.

Ans to Q No 4: When population is large, it becomes difficult to collect data from each unit of the population. In this case, cluster sampling is used. Cluster sampling is based on the ability of the researcher to divide the sampling population into groups called clusters, and then to select elements within each cluster, using the simple random sampling technique.



2.10 MODEL QUESTIONS

Q 1: What is simple random sampling? Explain with examples.

Q 2: State the advantages and disadvantages of simple random sampling.

Q 3: What are strata? Give examples.

Q 4: Explain stratified sampling and state the advantages and disadvantages.

Q 5: Explain systematic sampling with a suitable example.

Q 6: What is cluster sampling? Explain in details with examples.

*** ***** ***

UNIT 3: SAMPLING TECHNIQUES - II

UNIT STRUCTURE

3.1 Learning Objectives

3.2 Introduction

3.3 Non-Random Sampling Methods

- 3.3.1 Convenience Sampling
- 3.3.2 Purposive Sampling
- 3.3.3 Quota Sampling
- 3.3.4 Snowball Sampling
- 3.3.5 Volunteer Sampling

3.4 Let Us Sum Up

3.5 Further Reading

3.6 Answer to Check Your Progress

3.7 Model Questions

3.1 LEARNING OBJECTIVES

After going through this unit, you will be able to-

- differentiate between probability and non-probability sampling
- discuss the different types of non-probability sampling techniques
- discuss the advantages and disadvantages of non-probability sampling techniques.

3.2 INTRODUCTION

In our previous unit, firstly, we learnt about the concepts of sample and population and the differences between them. Secondly, types of sampling (probability and non-probability sampling techniques), particularly probability sampling techniques, were discussed in detail. In this unit, we will be learning about the non-probability sampling methods.

Non-probability sampling is a sampling procedure that does not afford any basis for estimating the probability that each item in the population has of being included in the sample. Non-probability sampling is also known by

different names such as deliberate sampling, purposive sampling and judgment sampling. In this type of sampling, items for the sample are selected deliberately by the researcher; their choice concerning the items remains supreme. In other words, under non-probability sampling, the organizers of the inquiry purposively choose particular units of the universe for constituting a sample on the basis that the small mass that they so select out of a huge one will be typical or representative of the whole. For instance, if a researcher wants to study the economic conditions of people living in a state, a few towns and villages may be purposively selected for intensive study on the principle that they can represent the entire state. The judgment of the organizers of the study plays an important part in this sampling design.

The prime difference between probability and non-probability sampling is that in the former type, each of the units of a population has the equal opportunity to get selected in a sample, whereas in the latter type, the sample is drawn based on some condition or willingness of the researcher. Hence all the units of the population do not have an equal chance of being included in the sample drawn.

3.3 TYPES OF NON-RANDOM SAMPLING METHOD

There are five types of non-random sampling or non-probability sampling methods-

3.3.1 Convenience Sampling

It is also known as 'accidental' sampling. It depends upon the convenience of the researcher in finding the sample. As such, the researcher considers the people readily found in that instant for his/her sample. For example- if someone wants to study the use of library among college students, he/she would consider the number of readers in the library; based on the attendance he/she would draw the conclusion. This method of sampling is common among market research and newspaper reporters. Especially during

elections, the media personnel take the opinions of the general public present in and around them. Again, approaching people standing in banks, grocery shops, malls, etc. and asking questions would also be an example of convenience sampling.

Advantages of convenience sampling include the easy access to sample population and it is economic in terms of money and time. Disadvantages of convenience sampling include the biasness of the researcher, since only a certain units of the population are taken into consideration for the study.

3.3.2 Purposive Sampling

It is also known as 'judgemental sampling'. Here, the researcher makes a judgement as to who would be the best one to include in the sample in order to get the desired responses. As such, he approaches only those people who can provide him with the information he/she is searching for. For example, the TV reporters stopping certain individuals on the street in order to ask their opinions about certain political changes constitutes the most popular example of this sampling method. However, it is important to specify that the TV reporter has to apply certain judgment when deciding who to stop on the street to ask questions; otherwise it would be the case of random sampling technique.

Let us consider another example. Suppose, a group of researchers wants to study the factors which lead a person to select ethical hacking as a profession. We know that ethical hacking is a skill which has been recently attracting youth. More and more people are selecting it as a profession. The researchers who understand what ethical hacking is will be able to decide who should form the sample to learn about it as a profession. That is when judgmental sampling is implemented. Researchers can easily filter out those participants who can be eligible to be a part of the research sample.

Economy in terms of money and time, avoidance of irrelevant units in the sample, better results when investigator is unbiased are

some of the advantages of purposive sampling. On the other hand, personal biasness and unequal chance of getting selected in the sample are some of the limitations of purposive sampling. Since selection of the sample is governed by some pre-determined purpose, hence sample selection at times is considered to be partial.



CHECK YOUR PROGRESS

Q 1: What is meant by non-probability sampling? Answer in about 50 words.

.....

.....

.....

.....

.....

Q 2: Differentiate between probability and non-probability sampling. Answer in about 60 words.

.....

.....

.....

.....

.....

Q 3: Mention the types of non-probability sampling. Answer in about 30 words.

.....

.....

.....

Q 4: Cite an example of convenience sampling. Answer in about 30 words.

.....

.....

.....

3.3.3 Quota Sampling

It is similar to stratified sampling with the only difference that in stratified sampling, population is divided into strata and units from each stratum are drawn randomly for the sample. Whereas, in quota sampling, there are certain fixed quotas based on which the sample is selected. The choice of the respondents is completely left to the interviewer. Let us consider the following example.

Suppose, a researcher wants to study certain behaviour of the college students of a local college. Because the researcher can access this data, s/he knows that in this given population, 43% of the students are male and 57% are female. For a sample size of 1,000, the researcher knows that 430 males and 570 females will need to be interviewed from that population.

Using the Quota Sampling Method: To more realistically represent a given population, it would be necessary to control more than one category. For example, the researcher could select gender and age, or gender and income, or level of education and income, depending on the purpose of the study. Using our example, the researcher could add age (less than 21 years and 21 years and over) as a control category, specifying that half of the sample for each gender be from each age category. Now we are searching for 215 males under 21 years and 215 males 21 years and over, along with 285 females under 21 years and 285 females 21 years and over. What has happened here is that with every added category, it may take longer to locate these individuals, thus adding cost and time to the process. However, the target population could be better represented if additional categories are considered.

Why use multiple quotas? The intent of specifying multiple control categories or quotas is to improve the representation of particular groups within the population and ensure that groups are not over-represented. In our example, males and females would be

interviewed until each category has reached its quota. Once 215 males under 21 years of age have been interviewed, no further male responses from that age category will be considered. Individuals for the remaining categories will continue to be interviewed until each quota is satisfied. This ensures that no more than 215 males and 285 females from each age category will be represented. In this way, no one category is over-represented according to the target population.

Advantages of quota sampling: Advantages of quota sampling are-(i) It is relatively less expensive (ii) Convenient (iii) Requires less time (iv) does not require sampling frames.

Disadvantages of quota sampling: Disadvantages of quota sampling are- (i) sample is not a true representative of the population (ii) presence of biasness of the interviewer (iii)amending results through a formal statistical way is not possible (iv) is not always fruitful in case of field works.



CHECK YOUR PROGRESS

Q 5: What is meant by quota sampling? Answer in about 40 words.

.....

.....

.....

3.3.4 Snowball Sampling

Snowball sampling is the process of selecting a sample using networks. It is also referred to as 'chain-referral' sampling. The researcher would first approach the respondents he/she knows and is available to him. As such, a few individuals in a group or organization are selected and the required information is collected from them. Then, these people provide references about other individuals they know who fall under the criteria of the researcher. Thus, they are asked to identify other people in the group or

organization, and the people selected by them become a part of the sample. Information is collected from them, and then these people are asked to identify other members of the group and, in turn, those identified become the basis of further data collection. This process is continued until the required numbers of respondents have been interviewed or no other respondents are discovered. In other words, till the saturation point has been reached, in terms of data collection, the process of snowball sampling continues.

For example, suppose we want to study about the work culture in foreign institutions. For this, firstly we would approach the ones who we know in and around us working in foreign land, they will then provide us the information about people they know who are associated with foreign institutions, from them again we are going to have further links to approach employees in foreign institutions. In this way, the process will continue till we find an adequate number of respondents. Again, if a researcher wants to study about the incidence of some rare disease, firstly he will collect information from the people he knows suffering from such disease and then from his/her link would be finding other such patients suffering from the disease.

Advantages of snowball sampling: This sampling technique is useful if we know little about the group or organization we wish to study, as we need only to make contact with a few individuals, who can then direct us to the other members of the group. This method of selecting a sample is useful for studying communication patterns, decision making or diffusion of knowledge within a group. Researchers can collect data from people who would have been otherwise difficult to approach. It is also economic in terms of cost and time, since it does not need to hire any extra research team for the purpose.

Disadvantages of snowball sampling: Firstly, the choice of the entire sample rests upon the choice of individuals at the first stage. If they belong to a particular faction or have strong biases,

the study may be biased. Secondly, it is difficult to use this technique when the sample becomes fairly large. Thirdly, in this method it is difficult to determine sampling error. Fourthly, here we are not really sure about the size of the population and hence fixing a sample size becomes difficult. Moreover, the sample chosen might not always be the true representative of the population.

3.3.5 Volunteer Sampling

In this method, the respondents himself/herself volunteers to give information he has. This implies respondents self-select to become a part of the sample. For example- if a researcher wants to study about the number of beneficiaries of a government policy and publishes it in newspapers with a contact number to send details, and people who have been benefited reply to it on their own, it would be an instance of volunteer sampling.

Advantages of volunteer sampling: Although this method is very rarely used, it has many advantages over the other methods. It is considered to be more ethical because respondents approach the researcher and provide information willingly. While doing so, generally they are not bias and speak the truth. Moreover, it is convenient and easy for the researcher since it involves no cost in approaching respondents for data collection.

Disadvantages of volunteer sampling: Sometimes, the sample chosen may not be representative of the population. As a result, there may be biasedness since the researcher makes little effort to control the sample composition, it includes only those respondents who voluntarily come forward. In this process most of the genuine respondents might be left from the sample.



CHECK YOUR PROGRESS

Q 8: How is the sample selected in purposive sampling? Answer in about 60 words.

.....

.....

.....

.....

.....

Q 7: Define Snowball sampling. Answer in about 50 words.

.....

.....

.....

.....

.....

Q 8: When does the process of sample selection end in case of Snowball sampling? Answer in about 50 words.

.....

.....

.....

.....

.....



3.4 LET US SUM UP

- Non-probability sampling is that sampling procedure which does not afford any basis for estimating the probability that each item in the population has of being included in the sample. Non-probability sampling is also known by different names such as deliberate sampling, purposive sampling and judgment sampling.
- The prime difference between probability and non-probability sampling is that in the former type each of the units of a population

has the equal opportunity to get selected in a sample, whereas in the latter type sample is drawn based on some condition or willingness of the researcher, hence all the units of the population do not have a equal chance of being included in the sample drawn.

- There are five types of non-random sampling or non-probability sampling methods- convenience sampling, purposive sampling, quota sampling, snowball sampling and volunteer sampling.
- Convenience sampling is also known as 'accidental' sampling. It depends upon the convenience of the researcher in finding the sample. As such, the researcher considers the people readily found in that instant for his/her sample.
- Purposive sampling is also known as 'judgemental sampling'. Here, the researcher makes a judgement as to who would be the best one to include in the sample in order to get the desired responses. As such, he approaches only those people who can provide him with the information he is searching for.
- Quota sampling is similar to stratified sampling with the only difference that in stratified sampling, population is divided into strata and units from each stratum are drawn randomly for the sample. Whereas, in quota sampling, there are certain fixed quotas based on which the sample is selected. The choice of the respondents is completely left to the interviewer.
- Snowball sampling is the process of selecting a sample using networks. It is also referred to as 'chain-referral' sampling. The researcher would first approach the respondents he/she knows and is available to him, from this initial group of respondents the researcher would be getting information about other respondents that matches with his criterion. This process of adding units to sample continues till an adequate number of units in the sample is found.
- In volunteer sampling, the respondents himself/herself volunteers to give information he has. This implies respondents self-select to become a part of the sample.



3.5 FURTHER READING

1) Ahuja, R. (2018). Research Methods. Jaipur: Rawat Publications.

2) Kothari, C.R. (2011). Research Methodology, Methods and Techniques. New Age International (P) Ltd.

3) Kumar, R. (2011). Research Methodology. New Delhi: SAGE Publications India Pvt. Ltd.



3.6 ANSWERS TO CHECK YOUR PROGRESS

Ans to Q No 1: Non-probability sampling is that sampling procedure which does not afford any basis for estimating the probability that each item in the population has of being included in the sample. Non-probability sampling is also known by different names such as deliberate sampling, purposive sampling and judgment sampling.

Ans to Q No 2: In probability sampling, each of the units of a population has the equal opportunity to get selected in a sample, whereas in non-probability sampling, the sample is drawn based on some condition or willingness of the researcher, hence all the units of the population do not have an equal chance of being included in the sample drawn.

Ans to Q No 3: There are five types of non-random sampling or non-probability sampling methods- convenience sampling, purposive sampling, quota sampling, snowball sampling and volunteer sampling.

Ans to Q No 4: If someone wants to study the use of library among college students, he/she would consider the number of readers in the library and based on the attendance he/she would draw the conclusion. Here the researcher approaches the readers available in the library

as per his convenience and accordingly considers them into the sample.

Ans to Q No 5: Quota sampling is similar to stratified sampling with the only difference that in stratified sampling, population is divided into strata and units from each stratum are drawn randomly for the sample. Whereas, in quota sampling, there are certain fixed quotas based on which the sample is selected.

Ans to Q No 6: In purposive sampling, the researcher makes a judgement as to who would be the best one to include in the sample in order to get the desired responses. As such, he approaches only those people who can provide him with the information he is searching for.

Ans to Q No 7: Snowball sampling is the process of selecting a sample using networks. It is also referred to as 'chain-referral' sampling. The researcher would first approach the respondents he/she knows and is available to him, from this initial group of respondents the researcher would be getting information about other respondents that matches with his criterion.

Ans to Q No 8: The process is continued until the required numbers of respondents have been interviewed or no other respondents are discovered. In other words, till the saturation point has been reached, in terms of data collection, the process of snowball sampling continues.

Ans to Q No 9: It is considered to be more ethical because respondents approach the researcher and provide information willingly. While doing so, generally they are not bias and speak the truth. Moreover, it is convenient and easy for the researcher since it involves no cost in approaching respondents for data collection.



3.7 MODEL QUESTIONS

Long Questions. (Answer each question in around 300-500 words)

Q 1: Highlight the prime differences between probability and non-

probability sampling techniques.

Q 2: Explain the types of non-probability sampling with appropriate examples.

Q 3: Explain the advantages and disadvantages of non-probability sampling techniques.

*** ***** ***

UNIT 4: SAMPLING TECHNIQUES - III

UNIT STRUCTURE

4.1 Learning Objectives

4.2 Introduction

4.3 Meaning and Definition of Variables in Research

4.4 Various Types of Variables

- 4.4.1 Quantitative Variables
- 4.4.2 Qualitative Variables
- 4.4.3 Independent and Dependent Variables
- 4.4.4 Extraneous Variables
- 4.4.5 Background Variables
- 4.4.6 Intervening Variables
- 4.4.7 Suppressor Variables
- 4.4.8 Moderating Variables

4.5 Population or Universe of a Study

- 4.5.1 Finite and Infinite Population
- 4.5.2 Homogeneous and Heterogeneous Population
- 4.5.3 Existent and Hypothetical Population
- 4.5.4 Known and Unknown Population

4.6 Sample as a Concept

- 4.6.1 Characteristics of an Ideal / Good Sample
- 4.6.2 Population and Sampling Methods
- 4.6.3 Difference between Census and Sampling Methods

4.7 Let Us Sum Up

4.8 Further Reading

4.9 Answer to Check Your Progress

4.10 Model Questions

4.1 LEARNING OBJECTIVES

After going through this unit, you will be able to-

- put forward the meaning of variables in research
- discuss different types of variables

- distinguish between the major types of variables
- discuss the concept of population of a study
- discuss the concept of sample.

4.2 INTRODUCTION

We all know that variables form an important part in our studies. Just in case of demand and supply, we have been explaining that they depend on a set of variables like quantity and price and many other related variables. In research, we shall have to deal with different types of variables. These variables may be broadly categories as dependent and independent, qualitative and quantitative variables, extraneous variables, background variable etc. This unit basically deals with variables. Here, the concept of variable has been discussed. Again, the concept of universe or population and sample have also been discussed.

4.3 MEANING AND DEFINITION OF VARIABLES IN RESEARCH

In order to comprehend research, variables are important. The researcher can get a sense of the important variables related to that particular study by reviewing the literature, and data can be acquired from primary or secondary sources. In the framework of a research study, the study's factors played a crucial role. These factors we can call as variables. A variable is something that has a value that changes over time. Some of the common examples of variables include age, religion, gender, level of education, income, consumption, expenditure, family size, country of birth, capital expenditure, height, blood pressure readings etc. The value of each of these variables differ from one individual to other. Thus, any attribute, characteristic, number, or quantity that changes over time or can take on different values in different conditions is referred to as a variable in research. Such examples are as follows: The type of product is the variable in this case if an investigator wants to compare the efficacy of four different items, A, B, C, and D. A market survey based on customer preferences can be

beneficial to a company researcher. The variable here is the consumer's preference. A social scientist could investigate the role of education in minimizing child marriages. The child marriage is the variable here. Similarly, we can add other factors to these examples, such as efficacy, effectiveness, education, gender, customer attitude, and so on. It depends on the study's goal and what a researcher is attempting to focus on.

4.4 VARIOUS TYPES OF VARIABLES

We can separate the variables in many ways based on various background characteristics:

- **Quantitative Variables:** are those variables which can be measured in numbers: These are of two types viz. (a) Continuous Variable and (b) Discrete Variable.
- **Qualitative Variables:** are those variables which can't be measured in numbers
- **Independent Variables and Dependent Variables:** Independent variables shows the impact on dependent variable.
- **Extraneous Variables:** These are not independent variable but may affect the results of estimation.
- **Background Variables:** These variables basically cover the demographic and socio-economic factors which tells us about the background characteristics of the study.
- **Intervening Variables:** These are the variables which help in understanding the relationship between other two variables.
- **Suppressor Variables:** These are the variables which don't have any relation or impact on the dependent variable but still help in predictive validity. This concept was first introduced by Horst in 1941.
- **Moderating Variables:** These are those variable which actually strengthen the association between variables.

Let's take a look at these variables in more detail:

4.4.1 Quantitative Variables

Quantitative variables, often known as numerical variables, are variables that may be measured numerically. We can use a variety of quantitative and statistical methods to measure these characteristics. The value of total production in each state would be a quantitative variable if a researcher wanted to measure the productivity levels of different states in India. Age, height, family size, income, consumption, and expenditure are examples of quantitative variables. That is, each of the aforementioned characteristics or features varies or differs from person to person. We call these variables quantitative or numeric variables since they are expressed in numbers. There are two types of quantitative variables: continuous and discrete variables. These variables are unpredictably distributed. Any variable with an undetermined value is known as a random variable. The following are the specifics of discrete and continuous variables:

- Continuous Variable: A continuous variable is one whose value may be determined through measurement. For example: height and weight of students in a class, time and distance etc.
- Discrete Variable: A discrete variable is one whose value may be calculated by counting. For example: number of students in a class, number of participants in a competitive exam, number of events in a program etc.

4.4.2 Qualitative Variables

Qualitative variables are those that can't be assessed directly, and there are various ways for measuring qualitative variables in research. The proxy variable can be used to describe the qualitative variable and this will give a way to measure the qualitative variable indirectly. If a researcher tries to give numerical ordering to a qualitative variable then that wouldn't be possible as it doesn't follow any logical ordering. The value of the variable

'gender' viz. male, female and transgender varies qualitatively; there is no ranking for any of these. A different terminology is also given to these kind of variables viz. categorical variables. As gender variable is expressed in categories, then this variable is considered as categorical variable. Similarly, place of residence is a categorical variable that can be defined as either rural or urban. Nominal and ordinal variables can be used to define categorical variables. The following is a description of ordinal and nominal variables:

- Ordinal variables, such as exam grades (O, A+, A, B+, and so on), can be logically ordered or rated higher or lower than another
- Nominal variables, for example, cannot be ranked or logically ordered. A qualitative variable is a characteristic that cannot be measured but can be classified as having or not having certain features. The example for this is type of gender, religion etc.

4.4.3 Independent and Dependent Variables

Many research efforts attempt to identify and understand the sources of the occurrences or problems, in order to create a causal relationship. Look at these statements to understand the relationships between variables and how we may determine which variables are dependent or independent:

- Inappropriate dietary choices causes malnutrition among children
- Alcohol enhances the risk of lung cancer.
- The drug causes the improvement of a health problem.

We have two variables in each of the above statements: one independent and one dependent. In the first instance, it is claimed that 'inappropriate dietary choices' created the 'malnutrition condition.' Therefore, it is considered as independent variable. Malnutrition is the dependent variable, because we assume that

this malnutrition was caused inappropriate dietary choices. Similarly, smoking and drug are the independent variables, and lung cancer, improvement in health are dependent variables. In general, an independent variable helps in measuring its effects on the dependent variable. The classification of both the variables is as follows:

A dependent variable is one that is utilised to characterise or evaluate the problem or outcome being studied. In a causal relationship, the cause is the independent variable, while the result is the dependent variable. The independent variable is smoking, and the dependent variable is cancer, if we think that smoking causes lung cancer. Depending on the context, a dependent variable can be referred to as a predicted variable, measured variable, explanatory variable, experimental variable, responding variable, outcome variable, output variable, or label.

The component that is considered as a cause or impact of the problem/result is described or quantified by an independent variable. The investigator attempts to describe or explain the impact or effect of the variable on the dependent variable. Variability in the dependent variable is presumed to depend on variability in the independent variable. An independent variable is sometimes called an explanatory variable, predictor variable, manipulated variable, controlled variable etc.

4.4.4 Extraneous Variable

Generally, in various research studies, the focus is on identifying one independent variable and measuring its impact on the dependent variable. However, various other variables alter our predicted independent-dependent variable relationship. Extraneous variables aren't always included in studies.

4.4.5 Background Variable

Different demographic and socio-economic factors provide a data package that tells us about the study's background features.

These variables are related to many independent factors, which influence or influence the dependent variable indirectly. As a consequence, they are referred to as background variables. If background components are important to the investigation, they should be measured.

4.4.6 Intervening Variable

A third variable is often responsible for an apparent link between two variables. Variables A and B, for example, may be highly correlated, but only because A causes the third variable, C, causing B in turn. C is the intermediary variable in this situation.

4.4.7 Suppressor Variable

In many circumstances, we have reason to suppose that the variables of interest are linked within themselves, yet our data contradicts this. It's likely that some unseen factors are concealing the true relationship between the two variables. Because it suppresses the true relationship between the other two variables, such a factor is known as a suppressor variable.

4.4.8 Moderating Variable

A simple one-to-one relationship must be modified to reflect other variables to properly characterise the link in the context of real-world research. This emphasises the significance to include a second indépendant variable (moderating variable), whose influence on the dependent and independent variable relationship is predicted to be significant.

4.4.9 Attributes

An attribute is a quality or characteristic of a specific variable. Variables and attributes are inextricably linked. A variable is a logical collection of characteristics. As variables can "vary,"

such as high and low scores. Thus these high and low scores are the attributes of a particular variables. The theory of attributes is concerned with qualitative characteristics that can be calculated using quantitative measurements. As a result, the attribute necessitates slightly different statistical treatments than the variables. Attributes are characteristics of the item under investigation, such as a smoking or drinking habit. As a result, the terms 'smoking' and 'drinking' both refer to an example of an attribute.



CHECK YOUR PROGRESS

Q 1: What is meant by background variables? Answer in about 50 words.

.....

.....

.....

.....

.....

Q 2: Define the term variable in the context of research. Answer in about 60 words.

.....

.....

.....

.....

.....

4.5 POPULATION OR UNIVERSE OF A STUDY

The term "universe" or "population" refers to a collection of items about which we obtain information. The term population refers to the entire field of study for which knowledge is sought. For example, if a research chooses to take all 1500 students studying in a college to assess the average height of the students, the full aggregate of 1500 students will be referred

to as the study's population. It is possible to have a finite or infinite population. Finite population is defined as a population with a finite number of things. On the other hand, infinte population refers to a population that has not been enumerated. The detailed description of finite and infinite population is as follows:

4.5.1 Finite and Infinite Population

Finite population refers to a population with a finite number of units that can be counted exactly. Here are a few examples of finite population:

- Population of the students who appeared in NEET Exam in India in September 2021.
- Students enrolled in Post Graduate courses in India.
- Number of teacher teaching in the University
- The population, in which the number of units is infinite and cannot be counted is called infinite population. The following are some examples infinite population:
- Decimal numbers lying between integers 1 and 2.
- Number of stars in the sky.
- Words formed by using all English alphabets.
- Number of cells in human body.

4.5.2 Homogeneous and Heterogeneous Population

A homogeneous population is one in which all of the units of the population are same or similar in terms of some characteristics. If all the units of population differ completely or in some aspects with one another, the population is called heterogeneous population. The mental and physical talents of all pupils in the same school and class differ from one another. This is an example of a population that is heterogeneous. Homogeneous is that where all the units possesses same characteristics.

4.5.3 Existent and Hypothetical Population

Existent population refers to population units that have a physical existence. In most circumstances, finite and existing populations can be thought of as alternatives to one another. A population of this type is students and teachers in a city, district, or state. Other examples of existing population are family members, working women in a metropolis, and children with autism or dyslexia in a certain area of the city. Hypothetical population refers to a population whose units do not have physical existence but whose existence is assumed or whose probability of existence is determined using statistical methods. This type of population is sometimes referred to as a statistical population. The probability of such a population is determined by the repetition of a previous incidence or by statistical computation. The following are some examples of such a population. A population depicting the average lifespan of electric lights manufactured in a plant. The bulbs' life expectancy is based on the durability of bulbs that have been made and utilised in the past. It is impossible to determine the exact life of such goods. Maximum and minimum temperature of different cities can be assumed on the basis of past experience. It cannot be predicted precisely.

4.5.4 Known and Unknown Population

If the parameter of the population is known, it is called known population. Parameter refers to statistical measurement taken from the data from the entire population. For example, in the study of achievement of the students appeared in S. S. C. examination in 2019-2020, the set of all the students appeared in SSC exam in that year will be called known population because the average marks of all these students and standard deviation of their marks can be calculated or known easily. If the parameter of population is unknown or cannot be calculated easily, it is known

as unknown population. For example, in the study of intelligence of secondary school students of Assam, the population will be of unknown type because average marks and standard deviation of scores of the students in intelligence test are not easily available and cannot be calculated easily because this population consists of a large number of students.

4.6 SAMPLE AS A CONCEPT

A sample is a subset of a population. In other words, a sample is a collection of selected or sorted units from a population. In other terms, a sample is a subset of the population that we select for research purposes. If a researcher chooses 400 students from a population of 1500, these 400 students are referred to as a sample. It means, the units, selected from the population as a sample, must represent all kind of characteristics of different types of units of population i.e. it should be representative.

4.6.1 Characteristics of an Ideal / Good Sample

Characteristics of an ideal sample are as follows.

- Number of units in sample must be proportionate. It means a size of sample must be in proper proportion of number of units in population.
- Units selected in sample must represent all the characteristics of different units of population.
- Sample should be helpful in realising all the objectives of research.
- Sample selection should be free from any kind of biases. It means all the units of population must have equal chance to be selected in the sample.
- Sample should be such, that can save time, energy and money of researcher.
- Contact of units of sample should be convenient for researcher. It means, units of sample

- should be within the reach of researcher.
- Collection of data from units of sample should be convenient for researcher. Selection of such sample makes the task of researcher easy and precise. Therefore sample is very much important in research.

Table 10.1 describe some terms related to sample:

4.1: Certain related Concepts in Sampling

Concept	Description
Sample	A representative group of individuals from a specific population
Sampling	The process of selecting a representative group of people or elements from a population. There are various methods of selecting a sample viz. probability and non-probability sampling methods.
Sampling Error	Sampling error is the difference between the means of sample groups and the population means. The term "sampling errors" refers to errors that occur as a result of the sampling procedure. It occurs as a result of poor sample method selection, erroneous sampling unit demarcation, population variability with various characteristics, and so on.
Non Sampling Errors	These errors are those which creep in due to human factors which always varies with one investigator to another. These errors general ling arise due to faulty planning, faulty selection of the sampling units, errors in compilation, negligence and non-response on the part of the respondent etc
Steps involved in the sampling process	(i)Define the population (ii)Identify the sampling frame. The sampling frame should be so selected which consists of almost all the sampling units. (iii) Specify sampling unit and sampling method (iv) Determine the sample size (v) Specify sampling plan (vi) Select the sample

4.6.2 Population and Sampling Methods

Any research can use a variety of data collection methods to obtain any measurement. Data can be gathered from either primary or secondary sources. A research study on the other hand, can use census methods or sampling methods to get an estimate on a specific research problem. The detailed description

of both the methods is as follows:

Census Method: The census technique is a method in which information or data is collected from each and every unit of the population in relation to the subject under inquiry, and conclusions are reached based on this information. Complete enumeration is another name for this procedure. The following are the merits and demerits of the Census method.

Merits of Census Method

- Data produced by the method of the census have greater reliability and precision as data from each unit of the universe are collected in this method.
- This method provides detailed information about every unit of the universe
- The strategy is better appropriate for a small population and a wide variety of attributes.

Demerits of Census Method

- The census approach is a costly one. As data is collected from each unit of the population, more money is required. This is why governments choose to employ this method for critical topics such as population censuses and agricultural censuses.
- This methods involves much time for data collection
- This type of data collection also necessitates a significant amount of labour. This necessitates a bigger number of enumerators.

Sampling Method: The sampling method is a data collection method in which data is collected from a sample of items drawn from a population and conclusions are drawn from them. In today's world, the sampling method is an important and widely used method of statistical investigation. In reality, there is almost no area where the sampling method is not used. The following are the merits and demerits of sampling methods.

Merits of Sampling Methods

- The sampling method is less costly. It saves time and money

because just a small percentage of the population is studied.

- Data can be collected more quickly using a sampling method because it is obtained from some items in the universe. As a result, this method saves time.
- As number of items is less in sampling method, they can be intensively studies
- In this method research work can be organized and executed more conveniently
- Sampling method is more scientific
- If the sample is selected in a way that entirely represents the universe, the outcome is more accurate and reliable.
- When it is impossible to conduct an investigation using the census method, the sampling method is more appropriate. In cases where the population is infinite or too large, the sampling method is used.

Demerits of Sampling Methods

- The sampling method is less accurate because, rather than enquiring about every unit in the universe, a partial inquiry or inquiry about a subset of units is conducted.
- If the method of sample selection is not unbiased or proper caution is not used, the results will undoubtedly be misleading.
- The investigator's bias is more likely, making the results less reliable.
- If all population units are different in that case, the sampling process will not prove very useful.

4.6.3 Difference between Census and Sampling Methods

On the basis of the above characteristics explained for both the methods, the main differences between the census and the sampling method are as shown in Table 10.2.

Table 4.2: Census Vs Sampling Methods

Basis of Comparison	Census Method	Sampling Method
Meaning	The census technique is a method in which information or data is collected from each and every unit of the population in relation to the subject under inquiry, and conclusions are reached based on this information.	The sampling method is a data collection method in which data is collected from a sample of items drawn from a population and conclusions are drawn from them.
Scope	All items related to the universe are investigated	Only a few items are inquired
Cost	Expensive	Economical
Field of investigation	Used for investigations with limited field	Used for investigations with large field
Results	Results would be more reliable and accurate	Results are less reliable and accurate due to various sampling and non-sampling errors during data collection
Enumeration	Complete	Partial
Homogeneity	This method is useful when units of the population are heterogeneous	This method is useful when units of the population are homogeneous
Type of universe	In such field where study of each and every unit of the universe is necessary, census method is more appropriate.	When population is infinite or vast, then sampling method is considered to be more appropriate.
Error	Not Present	Depends on the size of the population
Economy	It requires a large amount of money, time, and labour.	Relatively less money, time, and labour are required.
Organisation and Supervision	It is very difficult to organise and supervise the census method.	Easy to organize and supervise
Nature of Method	Old and not a very scientific method	New and Scientific



CHECK YOUR PROGRESS

Q 3: What is meant by unknown population?

Answer in about 60 words.

.....

.....

.....

.....

.....

.....

Q 4: Define the term sample. Answer in about 60 words.

.....

.....

.....

.....

.....

.....



4.7 LET US SUM UP

- In order to comprehend research, variables are important. The researcher can get a sense of the important variables related to that particular study by reviewing the literature, and data can be acquired from primary or secondary sources.
- Any attribute, characteristic, number, or quantity that changes over time or can take on different values in different conditions is referred to as a variable in research.
- Quantitative variables, often known as numerical variables, are variables that may be measured numerically. We can use a variety of quantitative and statistical methods to measure these characteristics.
- A continuous variable is one whose value may be determined through measurement. For example: height and weight of students in a class, time and distance etc.
- A discrete variable is one whose value may be calculated by counting. For example: number of students in a class, number of participants in a competitive exam, number of events in a program etc.
- Qualitative variables are those that can't be assessed directly, and there are various ways for measuring qualitative variables in research. The proxy variable can be used to describe the qualitative variable and this will give a way to measure the qualitative variable indirectly.
- A dependent variable is one that is utilised to characterise or evaluate the problem or outcome being studied. In a causal relationship, the

cause is the independent variable, while the result is the dependent variable.

- Depending on the context, a dependent variable can be referred to as a predicted variable, measured variable, explanatory variable, experimental variable, responding variable, outcome variable, output variable, or label.
- The component that is considered as a cause or impact of the problem/ result is described or quantified by an independent variable.
- The term "universe" or "population" refers to a collection of items about which we obtain information. The term population refers to the entire field of study for which knowledge is sought.
- Finite population refers to a population with a finite number of units that can be counted exactly
- If the parameter of the population is known, it is called known population. Parameter refers to statistical measurement taken from the data from the entire population.
- A sample is a subset of a population. In other words, a sample is a collection of selected or sorted units from a population. In other terms, a sample is a subset of the population that we select for research purposes.
- The sampling method is a data collection method in which data is collected from a sample of items drawn from a population and conclusions are drawn from them. In today's world, the sampling method is an important and widely used method of statistical investigation.



4.8 FURTHER READING

1) Guthrie, G. (2010) Basic Research Methods: An Entry to Social Science Research, Sage Publications, New Delhi, India.

2) Kumar, R. (2011) Research Methodology, Sage Publications, New Delhi, India.

3) Nachmias, D., and Nachmias C. (2014) Research Methods in Social Sciences, St. Martin's Press, New York.



4.9 ANSWERS TO CHECK YOUR PROGRESS

Ans to Q No 1: Different demographic and socio-economic factors provide a data package that tells us about the study's background features. These variables are related to many independent factors, which influence or influence the dependent variable indirectly. As a consequence, they are referred to as background variables. If background components are important to the investigation, they should be measured.

Ans to Q No 2: A variable is something that has a value that changes over time. Some of the common examples of variables include: age, religion, gender, level of education, income, consumption, expenditure, family size, height, blood pressure readings etc. The value of each of these variables differ from one individual to other. Thus, any attribute, characteristic, number, or quantity that changes over time or can take on different values in different conditions is referred to as a variable in research.

Ans to Q No 3: If the parameter of population is unknown or cannot be calculated easily, it is known as unknown population. For example, in the study of intelligence of secondary school students of Assam, the population will be of unknown type because average marks and standard deviation of scores of the students in intelligence test are not easily available and cannot be calculated easily because this population consists of a large number of students.

Ans to Q No 4: A sample is a subset of a population. In other words, a sample is a collection of selected or sorted units from a population. In other terms, a sample is a subset of the population that we select for research purposes. If a researcher chooses 400 students from a population of 1500, these 400 students are referred to as a sample.



4.10 MODEL QUESTIONS

Short Questions (Answer each question in around 150 words).

Q 1: Discuss the Meaning and Definition of Variables in Research.

Q 2: Discuss the various Types of Variables.

Q 3: Discuss the concept of universe or population.

Q 4: Briefly describe the concept of sample.

Q 5: Distinguish between census and sample.

Q 6: Write short notes on the following:

(a) Quantitative Variables (b) Qualitative Variables (c) Independent and Dependent Variables (d) Extraneous Variables (e) Background Variables (f) Intervening Variables (g) Suppressor Variables (h) Moderating Variables (i) Finite and Infinite Population (j) Homogeneous and Heterogeneous Population (k) Existent and Hypothetical Population (l) Known and Unknown Population.

Essay type Questions: (Answer each question in about 300-500 words each.)

Q 1: What is meant by variable? Discuss its the various. Draw the major distinctions between quantitative variables and qualitative Variables.

Q 2: Briefly describe the concept of sample. Mention the important characteristics of a good sample. Distinguish between census and sample.

*** ***** ***

UNIT 5: THEORY OF ESTIMATION

UNIT STRUCTURE

5.1 Learning Objectives

5.2 Introduction

5.3 Estimators and Estimates

5.4 Types of Estimation

5.5 Criteria for a Good Estimator

- 5.5.1 Unbiasedness
- 5.5.2 Consistency
- 5.5.3 Efficiency

5.6 Let Us Sum Up

5.7 Further Reading

5.8 Answers To Check Your Progress

5.9 Model Questions

5.1 LEARNING OBJECTIVES

After going through this unit, you will be able to-

- discuss how the theory of estimation is used to draw inference about population characteristies from the information provided by the sample
- distinguish an estimator from an estimate
- explain the difference between point estimation and interval estimation
- select an appropriate statstic to serve as a best estimator of a population parameter
- have knowledge about the propertices that a good estimator should possess
- put forward the concept of unbiasedness, consistency and efficiency of an estimator.

5.2 INTRODUCTION

This unit deals with the basic concepts of the theory of estimation. We have already discussed about the terms like population, sample, parameter and statistic etc in unit a of this book. So in the light of these knowledge we may define estimation as the technique of coming to a conclusion, regarding the magnitude of a parameter based on the information provided in the sample. As parameters reflect the characteristis of the population, whose values are unknown, our only way to estimate their values is to depend on the random sample from the same population. Obviously greater the size of the sample better will be the estimate of the parameter.

If we consider the case of a normal distribution with parameters μ and σ, whose values are unkonwn, there will always be an infinite number of functions of the sample values, called statistic, which we may propose as the estimates of the parameters. But the best estimate will be the one which will be the nearest to the true value of the parameter to be estimated.

In the subsequant sections of the unit, we shall discuss about a few related terms and concepts like estimators and estimates, types of estimation to develope a clear understanding of the topic, Also we are going to learn the concept of unbiasedness, consistency and efficiency with the help of examples.

5.3 ESTIMATORS AND ESTIMATES

In the method of estimation we are interested in finding an estimated formula of the estimate of the parameter, which is called the **'estimator'**. Again the numerical value of the parameter obained on the basis of the random sample using that estimator is called the **'estimate'**.

Thus estimator is the sample statistic used for obtaining an estimate of the parameter being estimated. For example, we know that the sample mean $\bar{x}$ is used to estimate mean μ, then $\bar{x}$ is an estimator. Again, an estimate refers to the actual sample result used to represent μ. If from a sample of size n = 36, the avarage age is 20 years, these sample mean $\bar{x}$ = 20 years is the estimate.

5.4 TYPES OF ESTIMATION

There are two types of estimation for a population parameter. They are,

(i) Point estimation

and (ii) Interval estimation.

In the first type of estimation, in the light of the sample, one may put forward a single value (i, e a point) as the estimate of the parameter. That is why it is called point estimation.

Sometimes instead of finding a point estimator, one may be interested in finding an interval within which the true value of the parameter lies. The method of construction of such an interval is known as interval estimation. To understand the difference betwen two types of estimation, think of an enumerator collecting data on the number of members per family, i.e, average family size in Assam. On the basis of his sample of say, 1000 familices, the enumerator may find the average family size may be taken as 4.2. This is point estimation.

On the otherhand, again, on the basis of the same sample, he may find that the average family size supposed to lie between 3.9 to 4.6, so the interval 3.9 to 4.6 is the interval estimate.

5.5 CRITERIA FOR A GOOD ESTIMATOR

It follows that an estimator must possess some desirable properties to be a good estimator. Such properties must meet the criterion that a good estimator is one whose sampling distribution is clustered around the true value of the parameter being estimated, than that of any other estimator, so a statistic can be regarded as a good estimator of a parameter if it satisfies the following properties.

(i) Unbiasedness

(ii) Consistency

(iii) Efficiency.

(iv) Sufficiency

We shall discuss about these properties one by one.

5.5.1 Unbiasedness

It is very likely that, the value of a statistic calculated from a given sample will be above or below the actual values of the population parameter of interest, because of the presence of sampling error. So what is desirable is that, the expected value or mean of the possible values of a statistic estimated from over all possible samples should be equal to the population parameter being estimated. If this is true then the statistic is called are unbiased estimator of the population parameter.

So the sample statistic $\bar{x}$ will be an unbiased estimator of the population parameter θ, if,

$$E(\bar{x}) = \theta$$

where, $E(\bar{x})$ is the mean of the sample statistic $\bar{x}$.

In the sampling distribution of sample mean and sample proportion, we get,

$E(\bar{x}) = \mu$ and $E(\bar{p}) = P$

Therefore, $\bar{x}$ and $\bar{p}$ are unbiased estimators of the corresponding population parameters μ and P respectively. Similarly, the difference $\bar{x}_1 - \bar{x}_2$, i.e. the difference between two sample means, $\overline{p_1} - \overline{p_2}$, the difference between two sample proportions are unbiased estimators of population parameters $\mu_1 - \mu_2$ and $P_1 - P_2$ respectivaly.

Again, if, $E(\bar{x}) \neq \theta$, then $\bar{x}$ is called a biased estimator of the parameter θ, and the amount of bias is, $E(\bar{x}) - \theta$,

Again, the sample mean square, not the sample variance is an unbiased estimators of the population variance σ^2.

i.e., $E(s'^2) = \sigma^2$

Where, $s'^2 = \frac{1}{n-1}\sum_{i-1}^{n}(x_i - \bar{x})^2$ is called the sample mean square, However, for large n, the sample variance, i. $s^2 = \frac{1}{n}\sum_{i-1}^{n}(x_i - \bar{x})^2$

becomes an unbiased estimator of the population variance σ^2.

It may be however mentioned that many biased estimators tend to become unbiased estimators asymptotically, because the bias vanishes as n increases.

An important point to be mentioned here is that an unbiased estimator may not be unique. To explain this let us consider the case of a poisson distribution with parameter λ.

As we know for Poisson distribution mean = λ = variance

Now, as discussed earlier, the sample mean is an unbiased estimator of the population mean.

i.e. $E(\bar{x}) = \lambda$

And sample mean square is an unbiased estimator of the population variance.

i.e. $E(s'^2) = \lambda$

So both sample mean and sample mean square are unbiased estimators of the population parameter λ.



CHECK YOUR PROGRESS

Q 1: State true or false.

(i) The estimating function of a parameter is the estimator. (T/F)

(ii) An estimate T_n is said to be an unbiased estimate of the parameter θ, if $E(T_n) > \theta$. (T/F)

(iii) In sample from a normal population, the sample mean is an unbiased estimate of the population mean θ. (T/F)

(iv) In a random sample from a normal population, the sample variance is an unbiased estimate of the population variance. (T/F)

Q 2: Distinguish between estimator and an estimate.

Q 3: What is the difference betwen point estimator and an interval estimator.

5.5.2 Consistency

The property of an estimator being consistent is derived from the law of large numbers which states that, as sample size increases, the sample mean $\bar{x}$ tends to the population parameter μ. similarly, a point estimator is consistent, if its value, say, $\bar{x}$ tend to become closer to the population parameter μ, as the sample size increases. So we can define a consistent estimator as follows :

Let $\hat{\theta}_n$ be an estimate of θ. based on a sample of size n, then $\hat{\theta}_n$ will be a consistent estimator of θ, if,

$$P\left\{\hat{\theta}_n \to \theta\right\} \to 1 \text{ as } n \to \infty$$

That is the probability that $\hat{\theta}$ approaches θ is 1 as n becomes larger and larger, i.e. higher the sample size, more concentrated should be the distribution be around θ.

For a symmetrical distribution, we know that both sample mean $\bar{x}$ and sample median x_{med} are unbiased estimators for a population mean μ. so, in case of a symmetrical distribution, consistency is the only desirable criterion to test for a good estimator. so, we see that the sample mean $\bar{x}$ is a consistent estimator of the population mean μ because $v(\bar{x}) = \frac{\sigma^2}{x}$ tends to zero for large n.

Again, the sample proportion $\frac{}{p}$ is a consistent estimator of the population proportion P since $v(\bar{p}) = \frac{\sigma^2}{n} \to 0$ as $n \to \infty$

As a matter of fact, it is seen, that consistency is the desirable property of an estimator when the sample size is large. It is of no pratical signifcance when the sample is small.

Remark :

(i) Consistency and unbiasedness of an estimator doesnot necessarily go together, i.e an unbiased estimator need not be consistent and a consistent estimator may be biased.

(ii) Consistent estimators are not unique.

1. If $x_1, x_2 ...., x_n$ is a random sample from a Normal population $N(\mu_1 1)$, show that $t = \frac{1}{n}\sum_{i=1}^{n} x_i^{\ 2}$, is an unbiased estimator of $\mu^2 + 1$.

Solution: Since $x_1, x_2 ...., x_n$ is the sample from $N(\mu_1 1)$

$$\therefore E(x_1) = \mu,\ V(x_1) = 1 \quad \forall i = 1,2,..n$$

Now, $V(X_i) = E(X_i^2) - \{E(x_i)\}^2$

$$\Rightarrow E(X_i^2) = V(X_i) + \{E(X_i)\}^2$$

$$= 1 + \mu^2$$

Now, E(t) $$= E\left(\frac{1}{n}\sum_{i=1}^{n} x_i^2\right)^2$$

$$= \frac{1}{n}\sum_{i=1}^{n} E(x_i^2)$$

$$= \frac{1}{n}\sum_{i=1}^{n} (1+\mu^2)$$

$$= (1+\mu^2)$$

2. Prove that in sampling from a $N(\mu, \delta^2)$ population, the sample mean is a consistent estimator of μ.

Solution: We know that, in sampling from $N(\mu, \delta^2)$ population, the sample mean $\bar{x}$ is also normally distributed with mean μ and variance $\frac{\delta^2}{n}$.

$$\bar{x} \sim N\left(\mu, \frac{\delta^2}{n}\right)$$

Now, $\frac{\lim}{n \to \alpha} V(\bar{x}) = \frac{\lim}{n \to \alpha} \frac{\delta^2}{n} = 0$

$\therefore \bar{x}$ is consistent estimator of μ.

5.5.3 Efficiency

The efficiency f an estimator is measured in terms of it's variance : for the same population, if there are two consistent estimators for a parameter, then that consistent estimator with smaller variance is said to be more efficient, because it provides an estimate closer to the population parameter. Thus for a random sample of size n, if $\bar{\theta}_1$ and $\bar{\theta}_2$ are two consistent estimators of the population parameter θ, then the relative efficency of $\bar{\theta}_2$ compared

to $\bar{\theta}_1$ is given by.

$$E = \frac{v(\bar{\theta}_1)}{v(\bar{\theta}_2)}$$

obviously, E < 1.

Where, $\bar{\theta}_1$ is the most efficient estimator.

By most efficient estimator, we mean that estimator in a class of consistent estimator for a parameter, whose variance is less than that of any such estimator.

Illustrated Example 1. In a normal sample from a $N(\mu, \sigma^2)$ population the sample mean $\bar{x}$ is the most efficient estimator for estimating.

μ and $V(\bar{x}) = \frac{\sigma^2}{n}$

Let x_{med} be the median for this sample.

then $V(x_{med}) = \frac{\pi\sigma^2}{2n}$

Efficency of $x_{med} = \frac{V(\bar{x})}{V(x_{med})} = \frac{\sigma^2/n}{\pi\sigma^2/2x} = \frac{2}{\pi}$

= 0.64

Again, whenever we deal with unbiased estimators of a parameter, we prefer that estimator which has minimum variance. This leads us to the minimum variance unbiased estimator of θ.

Remark : The correlation co-efficient between a most efficient estimator and any other estimator with efficiency e is $\sqrt{e}$.

Sufficicncy: An estimator is said to be sufficient for a parameter, if it contains all the information in the sample regarding the parameter. That is sufficiency required that an estimator uses the entire sample informations in estimating the parameter of our interest.

For example the sample statistic $\bar{x}$ to be a sufficient estimator of the population parameter, say, μ, it must contain the maximum information and furnish as much of it as possible it obtaining the value of the parameter μ. A sufficient estimator, if available is always preferred over a non-sufficient estimator, because it allows the use of comprehensive sample information in estimation, so if we say

that the sample mean $\bar{x}$ is sufficient estimator of μ, it implies that no other estimator, say x_{med} can convey more information in estimating the parameter μ.

Ex : In a normal population with mean μ and variance σ^2 we have,

(i) $\bar{x} = \frac{1}{n}\sum x_i$ is a sufficient estimator for μ, where σ^2 known,

(ii) $= \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2$ is sufficient estimator for σ^2.



CHECK YOUR PROGRESS

Q 4: State true or false.

(i) In sampling from a normal population, the sample mean is a consistent estimator of the population mean μ. (T/F)

(ii) A consistent estimator is unique. (T/F)

(iii) A consistent estimator T of parameter θ is the most efficient estimator iff V(T) < V (T'), where T' is any other consistent estimator of θ. (T/F)

(iv) If T_1 is the most efficient estimator of a parameter θ and T_2 be another estimator of θ, these the efficiency e of T_2 is, $e = \frac{V(T_2)}{V(T_1)}$

(T/F)

Q 5: Prove that, if $T'' = \frac{1}{2}(T + T')$, where T and T′ are both efficient estimators with variance V, then, $V(T'') = \frac{1}{2}V(1+p)$.

Where p is the correlation co-efficient between T and T′.

Q 6: Define consistent estimator. Do consistency and unbiasedness in estimators necessarily go together.



5.6 LET US SUM UP

- Estimation is the technique of coming to a decision about the value of the parameter depending on the imformation provided by the random

sample.

- Estimation may be of two types - point estimation and interval estimation.
- The estimating function of a parameter is known as estimator.
- The particular value of an estimator is called the estimate of the parameter.
- The criteria for a good estimator are :
  unbiasedness
  consistency
  efficiency and sufficiancy
- An estimator T is said to be unbiased estimator of the parameter θ, if, E(T) = θ.
- θ_n is a consistent estimator of θ if θ_n converges to θ in probability for large n, where n is the sample size.
- An unbiased estimator need not be consistent and a consistent estimator may be biased.
- The sample mean is an unbiased estimator of the population mean.
- Sample variance is not an unbiased estimator of the population variance of
- If θ_1 is the most efficient estimator with variauce V_1 and θ_2 is any other consistent estimator with variauce V_2 then efficiency of θ_2 is.

  $e = \frac{V_1}{V_2}$, e < 1.
- The correlation co-efficient between a most efficient estimator and any other estimator with efficiency e is $\sqrt{e}$.
- t is called minimum variance unbiased estimor of θ, if.
  (i) $E(t) = \theta$ (ii) $V(t) < V(t')$
  where t/ is any other unbiased estimator of θ.



5.7 FURTHER READING

1) Gupta, S. C. and Kapoor, V. K (2002) Fundementals of Mathematical Statistics, Sulten Chand and sons.

2) Goon, A. M, Gupta, M. (1985). An Outline of Statistical Theory, Vol I, New World Press Pvt. Ltd.

3) Mukhopadhyay, P (1996), Mathematical Statistics, New Central Book Agency (P) Ltd.

4) Rao, C. R. (1963), Linear Statistical Inference and its Application, John Willey and sons.



5.8 ANSWERS TO CHECK YOUR PROGRESS

Ans to Q No 1: (i) True, (ii) False, (iii) True, (iv) False.

Ans to Q No 2: The estimator is the estimated formula of a parameter. Again, the numerical value of an estimator from a fixed set of values of a random sample is known as estimate. Thus estimate stands for the value of the parameter. For example, the sample mean $\bar{x}$ is an is an estimate of population mean μ and sample mean square s^2 is an estimate of the population variance σ^2, whereas the formula for $\bar{x}$ and s^2 are the estimators.

Ans to Q No 3: The point estimation comprises of finding out an estimated formula of a parameter and finding out the numerical value on the basis of a sample using that estimated formula.

On the otherhand, interval estimation is the method of finding an interval which covers the true parameter with an assigned probability, say $100(1-\alpha)\%$. Such an interval is called a confidence interval with confidence co-efficient $(1-\alpha)$.

Ans to Q No 4: (i) True, (ii) False, (iii) True, (iv) False.

Ans to Q No 5: $V(T^{//}) = V\left[\frac{1}{2}(T+T^{/})\right]$

$$= \frac{1}{4}V(T+T^{/})$$

$$= \frac{1}{4}\left[V(T)+V(T^{/})+2\,\mathrm{cov}(T.T^{/})\right]$$

$$= \frac{1}{4}[V + V + 2\rho V] \qquad \Theta\ V(T) = V(T') = V$$

$$\rho = \frac{\text{cov}(T, T')}{\sqrt{v(T)}\sqrt{v(T')}}$$

$$\therefore \quad \text{cov}(T, T') = \rho V$$

$$\therefore \quad v(T'') = \frac{1}{4}[2v(1+\rho)]$$

$$= \frac{1}{2}[v(1+f)]$$

Ans to Q No 6: Let θ_n be an estimator of θ based on a sample of size n, then θ_n is a consistent estimator of θ, if θ_n converges in probability to θ, for. large n, i.e.

$\theta_n \xrightarrow{p} \theta$ as $n \to \alpha$

Again, consistency and unbiasedness of an estimator doesnot necessarily go together. An unbiased estimator neednot necessarily be consistent. similarly, a consistent estimator may be biased.



5.9 MODEL QUESTIONS

Q 1: Define an estimator and an estimate. State the requirements of a good estimator.

Q 2: Define the following terms and give one example for each,

(i) Consistency,

(ii) Unbiasedness

(iii) Efficiency.

Q 3: Show that the unbiased estimators are not unique with the help of an example.

*** ***** ***

UNIT 6: TESTING OF HYPOTHESIS

UNIT STRUCTURE

6.1 Learning Objectives
6.2 Introduction
6.3 Statistical Hypothesis: Null Hypothesis and Alternative Hypothesis
6.4 Errors in Hypothesis Testing, Level of Significance and Critical Region
6.5 One -Tailed and Two-Tailed Tests
6.6 General Procedure for Hypothesis Testing
6.7 Testing of Hypothesis in case of Large Samples
- 6.7.1 Hypothesis Testing for Single Population Mean
- 6.7.2 Hypothesis Testing for Single Population Proportion
- 6.7.3 Hypothesis Testing for Difference between Two Population Means
- 6.7.4 Hypothesis Testing for Difference between Two Population Means

6.8 Let Us Sum Up
6.9 Further Reading
6.10 Answers To Check Your Progress
6.11 Model Questions

6.1 LEARNING OBJECTIVES

After going through this unit you will be able to

- establish null and alternative hypotheses about a population parameter
- understand the two types of errors; interpret the confidence level and the significance level
- develop hypothesis testing procedures for accepting or rejecting a null hypothesis
- discuss how and when to use the normal and t distributions for hypothesis testing.

6.2 INTRODUCTION

In today's dynamic business environment, decision makers need might answers to certain questions in order to take optimum decisions. For example a pharmaceutical concern might be interested in assessing whether a new drug is really effective in curing a disease; a personnel manager might be interested to study the job satisfaction level of employees, a marketing manager might be interested in knowing the customer satisfaction about a particular product, etc. In all the situations, the concerned manager has to take wise decisions since such decisions involve the risk of taking wrong decisions. It is here that modern theory of probability plays a vital role in decision making and the branch of Statistics that develops the criterion for such decisions is known as testing of hypothesis. Hypothesis testing begins with an assumption commonly termed as "hypothesis" that a researcher makes about an unknown population parameter. Hypothesis testing is a well-defined procedure which enables us to decide objectively whether to accept or reject the hypothesis based on the information obtained from the sample.

6.3 STATISTICAL HYPOTHESIS: NULL HYPOTHESIS AND ALTERNATIVE HYPOTHESIS

A statistical hypothesis is some assumption or statement, which may or may not be true, about an unknown population parameter. For example, suppose the Vice President (HR) of a company wants to examine the effectiveness of a training programme which the company has organized for all its employees located at different places of the country. The Vice President (HR) will then make an assumption that "training has not enhanced efficiency". To test such claim statistically, sample data are collected and analyzed. The assumption is accepted or rejected on the basis of the sample findings.

There are two types of statistical hypothesis: (i) Null hypothesis and (ii) Alternative hypothesis. In hypothesis testing, to reach decisions we must

start with a hypothesis, called null hypothesis which is symbolized H_0. The null hypothesis asserts that there exists no (significant) difference between the sample statistic and the population parameter and whatever observed difference is there is merely due to fluctuations of sampling from the same population. In the words of R.A. Fisher, "Null hypothesis is the hypothesis which is tested for possible rejection under the assumption that it is true." The null hypothesis that makes a claim regarding the specific value of the parameter is generally expressed by any one of the following mathematical statements

(i) $H_0 : \mu = \mu_0$ (ii) $H_0 : \mu \leq \mu_0$ (iii) $H_0 : \mu \geq \mu_0$

where μ is the population mean and μ_0 denotes a hypothesized value of μ.

An alternative hypothesis, generally denoted by H_1, is the opposite of the null hypothesis. Thus any hypothesis which is complementary to the null hypothesis is called an alternative hypothesis. It is obvious that both null and alternative hypotheses cannot be true simultaneously at the same time. In other words, when null hypothesis is found to be true, the alternative hypothesis is to be rejected or when null hypothesis is rejected, the alternative hypothesis must be true. For example, if we are interested in studying the hypothesis that the population has a specified mean, μ_0 (say), then we have the null hypothesis $H_0 : \mu = \mu_0$ and against this null hypothesis we may have any one of the following alternative hypotheses (depending upon the situation) given by,

(i) $H_1 : \mu \neq \mu_0 \quad (i.e., \mu > \mu_0 \text{ or } \mu < \mu_0)$

(ii) $H_1 : \mu > \mu_0$ (iii) $H_1 : \mu < \mu_0$

The alternative hypothesis given by (i) is called two-tailed alternative and the alternatives given by (ii) and (iii) are called right-tailed and left-tailed alternatives respectively. Accordingly, the nature of the alternative hypothesis enables us to decide whether we have to use an one-tailed (right or left) or two-tailed test.

Simple and Composite hypothesis: If the hypothesis completely specifies the population, then it is called simple hypothesis, otherwise it is called composite hypothesis. For example, from a normal

population $N(\mu, \sigma^2)$, the hypothesis of the form

$H: \mu = \mu_0$ and $\sigma^2 = \sigma_0^2$ (both μ and σ^2 are specified)

is a simple hypothesis because it completely specifies the distribution.

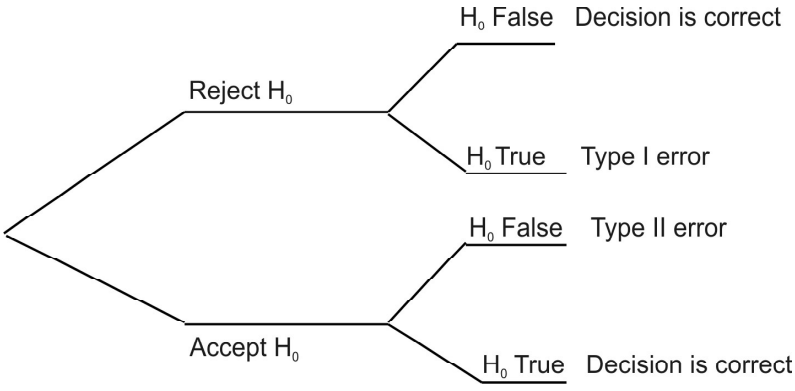
On the other hand, the hypotheses of the forms

(i) $H: \mu = \mu_0$ (σ^2 is not specified), (ii) $H: \sigma^2 = \sigma_0^2$ (μ not being specified) are composite since each of these does not completely specify the value of the parameter.

6.4 ERRORS IN HYPOTHESIS TESTING, LEVEL OF SIGNIFICANCE AND CRITICAL REGION

In hypothesis testing, the null hypothesis H_0 should be accepted if it happens to be true and the null hypothesis H_0 should be rejected if it happens to be false. However, in practical situations, the correct decision is not always possible while testing a null hypothesis. Since the decision to reject or accept a null hypothesis depends only on sample data therefore a decision maker may commit two types of errors (i.e., take wrong decisions) which are diagrammatically shown in Fig.13.1.

Figure 6.1: Type I and Type II errors



Type I error is committed if we reject a null hypothesis when it is true. The possibility of committing Type I error is denoted by the symbol α. Thus,

$\alpha = P$ (Type I error)= P (Reject H_0 when H_0 is true)

On the other hand, Type II error is committed if we accept a null hypothesis

when it is false. The possibility of committing Type II error is denoted by the symbol β. Thus we have

$\beta = P$ (Type II error)=(Accept H_0 when H_0 is false)

where α and β are also known as the sizes of Type I and Type II error respectively.

For example, a business firm, before introducing a new product, has to choose one of the two decisions- to introduce or not to introduce the product. Again if the product is introduced then the firm may face two situations- the failure of the product and the success of the product. The firm thus runs the risk of taking wrong decisions in two ways. The one risk of taking wrong decision known as Type II error, is that the firm does not introduce the product though it would have succeeded had it been introduced. The another risk of taking wrong decision known as Type I error is that the firm introduces the product but it failed.

In the terminology of Statistical Quality Control, Type I error is known as producer's risk while Type II error is known as consumer's risk.

In practice, it is not possible to make error free judgment in Statistics. Although both the errors should be very small yet it is not possible to minimize both these errors simultaneously at the same time on the same hypothesis test. Generally there exists an inverse relationship between α and β as a result of which an attempt to reduce type I error (α) leads to an increase in Type II error (β) and vice-versa. Since in most of decision making problems, it is more risky to accept a wrong hypothesis than to reject a correct one, i.e., consequences of Type II error are likely to be more serious than the consequences of Type I error therefore a compromise is made by minimizing the more serious error (i.e., β) after keeping fixed the less serious error (i.e., α).

In testing of hypothesis, the **Level of significance** is defined in terms of the probability of rejecting a null hypothesis when it is true. The level of significance is also known as the size of the critical region or the rejection region and generally denoted by α. It is to be noted that the level of significance should be determined before a test procedure so that obtained result is free from the influence of the decision maker. In practice, generally

adopted level of significance by researchers is: 1 % (=0.01), 5% (=0.05) and 10% (=0.10). If a researcher use 5% level of significance, it would mean that in 5 samples out of 100 he is likely to reject a correct null hypothesis H_0. In other words the researcher is 95% confident that his decision to reject H_0 is correct. Thus higher the significance level adopted for testing a hypothesis, the higher the chance of rejecting a true null hypothesis.

In hypothesis testing, the area under the sampling distribution curve of the test statistic is divided into two mutually exclusive regions termed as acceptance region and rejection or critical region. The acceptance region is a range of values of the sample statistic that lead us to the acceptance of null hypothesis H_0 while critical region is the range of sample statistic values within which if values of the sample statistic falls lead to the rejection of null hypothesis H_0. In Figure 13.2 if the calculated value of the sample statistic falls in the acceptance region, the null hypothesis is accepted otherwise it is rejected. For making a decision whether to accept or reject a null hypothesis, a researcher has to determine the critical value that separates the regions of acceptance and rejection. The determination of critical value depends on the size of the rejection region, which is directly associated with the risk involved in decision making.

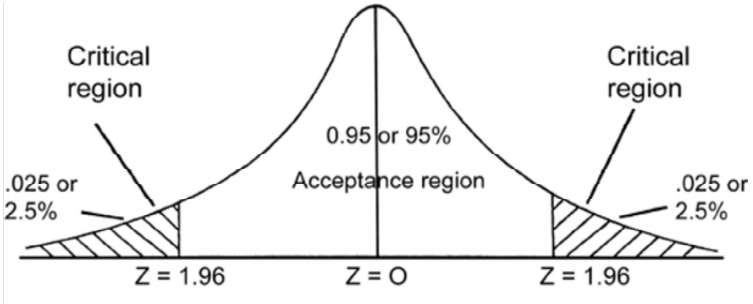
6.5 ONE TAILED AND TWO TAILED TESTS

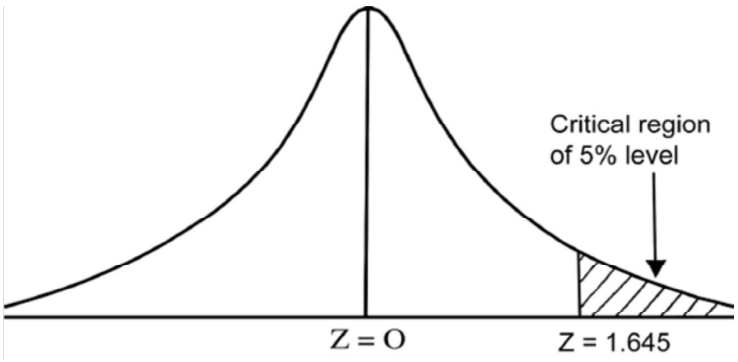
In any test, there are two types of tests of hypothesis. They are: (i) One-tailed tests and (ii) Two-tailed tests. Whether to use a one-tailed test or two tailed test depends entirely on the formulation of the alternative hypothesis. If the alternative hypothesis is one-tailed then the test to be applied is one-tailed and if the alternative hypothesis is two-tailed then the corresponding test is two-tailed. In case of two-tailed test, the critical or significant values of the test statistic lie towards both the tails of the graph of the sampling distribution of the statistic. Figure 13.2 relates to standard normal test statistic Z. The figure reveals that critical values of the standard normal test statistic Z. lies on both sides of the mean i.e., in both the tails of the distribution. The shaded region under the normal curve is the rejection region (or critical region) at 5% level of significance, the region being 5% of

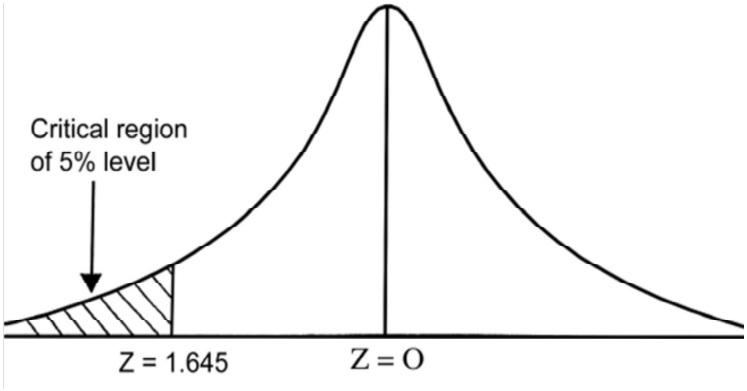
the total area. In other words, the total area under the normal curve being unity, the size of the rejection region is 0.05.

Again sometimes we may encounter certain situations where alternative hypothesis states that either one process is better than the other or one process is worse than the other. But in two-tailed test the alternative hypothesis H_1 states that both may occur i.e., one process may be better or worse than the other. Figure 13.3 and Figure 13.4 below exhibits the critical values and critical regions of the standard normal test statistic for right-tailed test and left-tailed test respectively for 5% level of significance.

Figure 6.2: Two-tailed Test







Now, we present below a summary of certain critical values of test statistic Z (i.e., standard normal test statistic) for both one-tailed and two-tailed tests at some commonly used level of significance. These critical values of Z for different levels of significance have been obtained by using the table of area under normal curve.

Table 6.1: Certain Critical Values of Test Statistic Z

Critical values (Z_α)	Level of significance (α)		
	5% or 0.05	1% or 0.01	10% or 0.010
Two-tailed test	$\|Z_\alpha\| = 1.96$	$\|Z_\alpha\| = 2.58$	$\|Z_\alpha\| = 1.645$
Right-tailed test	$Z_\alpha = 1.645$	$Z_\alpha = 2.33$	$Z_\alpha = 1.28$
Left-tailed test	$Z_\alpha = -1.645$	$Z_\alpha = -2.33$	$Z_\alpha = -1.28$

Note: (i) The notation Z_α stands for the critical value of Z at α level of significance.

(ii) Critical or significant values are also known as tabulated values.

(iii) If the size of the sample is small (usually less than 30) then the normality assumption of the sampling distribution of the test statistic is not valid and hence we cannot use the significant values given in the above table. In such situation we use the significant values based on the exact sampling distribution of Z, which turns out to be t, F or χ^2. (Please refer to any Standard textbook on Statistics).

6.6 GENERAL PROCEDURE FOR HYPOTHESIS TESTING

Hypothesis testing is a well-defined systematic procedure that involves five different steps as given below:

(1) The first step is to set up the null hypothesis H_0.

(2) Set up the alternative hypothesis H_1. The nature of alternative hypothesis H_1 helps us to choose whether to be adopted one-tailed or two-tailed test.

(3) To choose an appropriate level of significance. Commonly used levels of significance are: 0.01, 0.05 and 0.10.

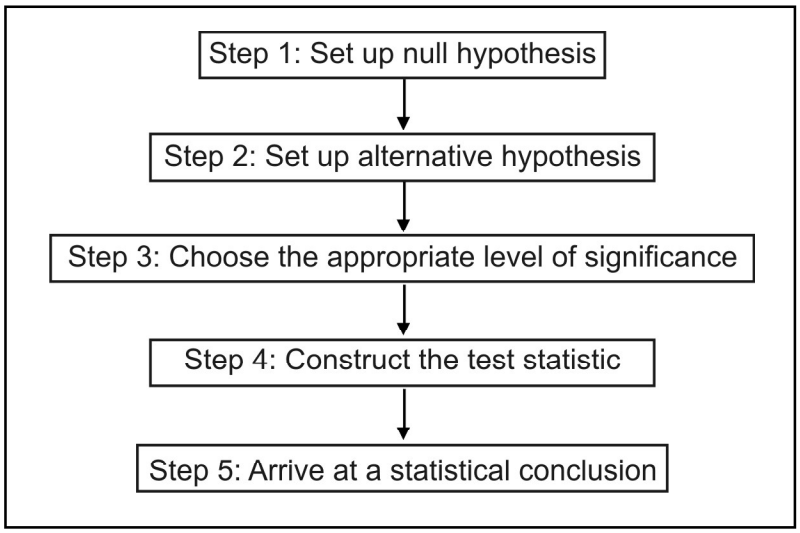
(4) After setting up the null and alternative hypotheses, the researcher

needs to select an appropriate statistical test that will be used for analyzing statistical data. Type, number and the level of data may provide a platform in the selection of the statistical test. Apart from these, the statistics used in the study (mean, proportion and variance etc.) must also be considered when a researcher selects an appropriate test, which can be used for testing of hypothesis to obtain the best results. Some of the widely used testing procedures are: z, t, F and χ^2.

(5) The final step in any hypothesis testing is to draw a statistical decision, which involves either to accept or to reject a null hypothesis. This depends on whether the calculated value of the test statistic falls in the acceptance region or in the rejection region at a given level of significance. In other words, if the computed value of Z (in modulus value) is less than the critical value Z_α then we say that it is not significant and null hypothesis may be accepted. On the other hand, if the computed value of Z (in modulus value) is greater than the critical value Z_α then we say that it is significant and null hypothesis is rejected at the level of significance α.

Therefore, various steps to be followed in testing of a statistical hypothesis can be summarized as shown in Figure 13.5

Figure 6.5: Steps in testing of a Statistical Hypothesis



Power of Test: Another way to evaluate the goodness of a statistical test is the complement of type II error, known as power of a test, which is defined as:

Power of a test = P (rejecting H_0 when it is false)

= P (rejecting H_0 when H_1 is true)

= 1- P (accepting H_0 when H_1 is true)

= 1- P (type II error)

= 1- β

Where β is the probability of type II error. Among a class of tests, the best is the one which has the maximum power for the same size.



CHECK YOUR PROGRESS

Q 1: What is meant by Hypothesis testing?

.....

.....

Q 2: What is neant by Type I error?

.....

.....

Q 3: What is meant by Type II error?

.....

.....

6.7 TESTING OF HYPOTHESIS IN CASE OF LARGE SAMPLES

Here we shall develop the tests of significance in case of large samples. In practice, a sample is to be considered as large if the sample size is greater than or equal to 30 i.e., $n \geq 30$. Now since the assumptions made in case of large samples are not valid for small samples therefore the tests of significance used to deal with problems relating to large sample are different from the ones used for small samples.

The following are the assumptions made in testing of hypothesis relating to large samples:

(i) The population from which the sample is drawn is normally distributed. Consequently the sampling distribution of the sample statistic is also normal.

(ii) Since sample size is large, due to central limit theorem, the value given by any random sample can be used in place of population value for calculating the standard error of the estimate.

6.7.1 Hypothesis Testing for Single Population Mean

Suppose a random sample of size $n(n \geq 30)$ with mean $\bar{x}$ is drawn from a population. We hold the claim that the population mean μ has the specified value μ_0.

Here we set up the null hypothesis $H_0 : \mu = \mu_0$. Against this hypothesis we may have any one of the following three alternative hypotheses

(i) $H_1 : \mu \neq \mu_0$ (Two-tailed test)

(ii) $H_1 : \mu > \mu_0$ (Right-tailed test)

iii) $H_1 : \mu < \mu_0$ (Left-tailed test)

Under the null hypothesis the test statistic to be constructed is:

$$Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

where $\mu = \mu_0$, σ is the population standard deviation, n is the sample size and $\bar{x}$ is the sample mean.

In many real life situations, the population standard deviation remains unknown. To overcome such type of situations, in case of large sample ($n \geq 30$), the unknown population standard deviation σ can be replaced by a sample standard deviation (s).

For the comparison of the calculated value of Z with the critical value at the given level of significance, Table 13.1 is used and on the basis of the comparison decision to accept or reject the null hypothesis is taken as per the procedure discussed in 13.6.

We will now discuss the following examples.

Example 1: A business firm conducted a sample survey 15 years ago and claimed that the average household income of a particular geographic region is Rs. 10,000. Mr. Baruah, recently appointed as vice president in the firm has expressed doubts about the accuracy of the data. In order to verify the claim, a random sample of 200 households is tested. It is found that the mean household income is Rs. 11,000. Assume that the population standard deviation of the household income is Rs. 1200. Can we support the claim at 0.05 level of significance?

Solution: Here first we set up the null hypothesis H_0 and alternative hypothesis H_1 as follows:

$$H_0 : \mu = 10{,}000$$

$$H_1 : \mu \neq 10{,}000$$

Given, $n = 200$, $\bar{x} = 11{,}000$, $\sigma = 1200$ and $\alpha = 0.05$

Since the sample size is large (≥ 30), therefore the Z test can be applied to test hypothesis.

Under the null hypothesis, the test statistic is given by

$$Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

$$= \frac{11{,}000 - 10{,}000}{\frac{1200}{\sqrt{200}}} = 11.79$$

Therefore the calculated value of Z is 11.79

Again the critical value of Z at 5% level of significance for two-tailed test =1.96 Since the calculated value is Z greater than the critical value of Z, therefore the statistical decision is to reject the null hypothesis and conclude that the claim is not valid. Mr. Baruah's doubt about average household income was right.

Example 2: A company conducted a survey in the past to its consumers which reveals that the average household income is Rs. 20,000 in a particular region. Due to the rapid growth of development the company thinks that the average income may have changed.

To verify this, a random sample of 50 households is selected and found that sample mean is Rs. 30,000 with a standard deviation of Rs. 15,000. Use 5% level of significance to verify the finding.

Solution: In the light of the problem, the null hypothesis H_0 and alternative hypothesis H_1 are given as

$$H_0 : \mu = 20{,}000$$

$$H_1 : \mu \neq 20{,}000$$

Given, $n = 50$, $\bar{x} = 30{,}000$, $s = 15{,}000$ and $\alpha = 0.05$

Since the sample size is large (≥ 30), therefore the Z test can be applied to test hypothesis.

Under the null hypothesis, the test statistic is given by

$$Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

$$= \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

(since the population S.D. σ is unknown, therefore it can be replaced by the sample S.D. s)

$$= \frac{30{,}000 - 20{,}000}{\frac{15{,}000}{\sqrt{50}}} = \frac{70.71}{15} = 4.714$$

Therefore the calculated value of Z is 4.714

Again the critical value of Z at 5% level of significance for two-tailed test =1.96 Since the calculated value Z is greater than the critical value of Z therefore the statistical decision is to reject the null hypothesis and conclude that average household income is significantly different from Rs. 20,000.

Example 3: The mean life of a sample of 80 fluorescent electric bulbs produced by a company is found to be 2000 hours with a

standard deviation of 400 hours. Test the hypothesis that the mean life time of the bulbs produced in general is higher than the mean life of 1900 hours at $\alpha = 0.01$ level of significance.

Solution: For the problem, let us take the null hypothesis that mean life of bulbs is not more than 1900 hours. Then the null hypothesis H_0 and alternative hypothesis H_1 will be as follows

$$H_0 : \mu \leq 1900$$

$$H_1 : \mu > 1900 \text{ (Right-tailed test)}$$

Given, $n = 80$, $\bar{x} = 2000$, $s = 400$ and $\alpha = 0.01$

Since the sample size is large (≥ 30), therefore the test can be applied to test hypothesis.

Under the null hypothesis, the test statistic is given by

$$Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

$$= \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

(since the population S.D. σ is unknown, therefore it can be replaced by the sample S.D. s)

$$= \frac{2000 - 1900}{\frac{400}{\sqrt{80}}}$$

$$= \frac{100}{44{,}72}$$

$$= 2.23$$

Therefore the calculated value of Z is 2.23

Again the critical value of Z at 1% level of significance for right-tailed test =2.33

Since the calculated value Z is less than the critical value of Z, therefore the statistical decision is to accept the null hypothesis.

Hence we conclude that mean life of bulbs produced by the company is not more than 1900 hours.

6.7.2 Hypothesis Testing for Single Population Proportion

There are many areas where information that is highly dynamic in nature is generally expressed in terms of proportions or percentage such as, we often find that the market share of a company is 40% or 25% of the customers have switched from one brand to another brand or 10% of the items are defective. Quality defects, consumer preferences, market share etc. are some of the common areas where researchers may be interested to check the hypothesis whether such proportions have changed. For this a random sample of size n is selected from a large population possessing a particular attribute of interest (also termed as success) X then

$$p = \text{observed sample proportion of successes} = \frac{\mathrm{X}}{\mathrm{n}}$$

Here we are interested to test the hypothesis that the proportion P in the population has a specified value P_0.

We set up the null hypothesis $H_0 : P = P_0$

Depending upon the nature, the alternative hypothesis may be any one of the following:

(i) $P \neq P_0$ (Two-tailed test)
(ii) $P > P_0$ (Right-tailed test)
(iii) $P < P_0$ (Left-tailed test)

Since n is large, therefore under the null hypothesis, the test statistic to be constructed is

$$\mathrm{Z} = \frac{\mathrm{p} - \mathrm{P}}{\sqrt{\frac{\mathrm{PQ}}{\mathrm{n}}}} \sim \mathrm{N}(0,1)$$

where $P = P_0, Q = 1 - P = 1 - P_0$

To compare the calculated value of Z with the critical value at the

given level of significance for the statistical decision, we use the Table 13.1.

Example 4: The production manager of a company that manufactures electric heaters claims that at least 10% of the heaters are defective. In order to test his believe, he takes a random sample of 100 heaters and finds that 12 heaters are defective. Use 5% level of significance to test the hypothesis.

Solution: Given $n = 100$, Number of defective heaters $= 12$

$p =$ sample proportion of defective heaters $= \frac{12}{100} = 0.12$

Null hypothesis $H_0 : P = 0.10$

Alternative hypothesis $H_1 : P \neq 0.10$

Since the sample size is large (≥ 30), therefore the Z test can be applied to test hypothesis.

Under the null hypothesis, the test statistic is given by

$$Z = \frac{p - P}{\sqrt{\frac{PQ}{n}}}$$

$$= \frac{0.12 - 0.10}{\sqrt{\frac{0.10 \times 0.90}{100}}} \quad (\Theta\ Q = 1 - P = 1 - 0.10 = 0.90)$$

$$= \frac{0.02}{0.03} = 0.67$$

$\therefore$ The calculated value of $Z = 0.67$

The critical value of Z for two-tailed test at 5% level of significance=1.96 Since the calculated value is less than the critical value (0.67<1.96) therefore the statistical decision leads to acceptance of null hypothesis and conclude that the production manager's claim that at least 10% of the products are defective is valid. Consequently, the manufacturer can plan a strategy considering the fact that 10% of the products may be defective.

Example 5: A manufacturer of LCD TV claims that it is becoming quite popular in Guwahati, and that about 5% homes are having LCD TV. However, a dealer of conventional TVs claims that the percentage of homes with LCD TV is less than 5%. A sample of 500 household is surveyed, and it is found that only 18 households have LCD TV. Test at 1% level of significance whether the claim of the company is tenable.

Solution: Given, Given $n = 400$, Number of households having LCD TV

$p =$ sample proportion of LCD TV owners $= \frac{18}{500} = 0.045$

Null hypothesis $H_0 : P = 0.05$

Alternative hypothesis $H_1 : P < 0.05$

Since the sample size is large (≥ 30), therefore the Z test can be applied to test hypothesis.

Under the null hypothesis, the test statistic is given by:

$$Z = \frac{p - P}{\sqrt{\frac{PQ}{n}}}$$

$$= \frac{0.045 - 0.05}{\sqrt{\frac{0.05 \times 0.95}{500}}} \quad (\Theta\ Q = 1 - P = 1 - 0.05 = 0.95)$$

$$= \frac{-0.005}{0.0097}$$

$$= -0.515$$

$$\therefore \quad |Z| = 0.515$$

Therefore the calculated value of $|Z| = 0.515$

The critical value of Z for one-tailed test at 1% level of significance = 2.33 Since the calculated value is less than the critical value (0.515 < 2.33) therefore the statistical decision leads to the acceptance of null hypothesis and conclude that the manufacturer's claim that about 5% homes in Guwahati are having LCD TV is valid.

Example 6: The manager of a workshop claimed that at least 95% of the equipments which he supplied to a factory conformed to

specifications. To verify his claim a sample of 200 pieces of equipments was selected and revealed that 18 were faulty. Test his claim at a significance level of 0.05.

Solution: Given n=200,

$P =$ Proportion of pieces conforming to specification in the lot = 95% =0.95

$Q =$ Proportion of pieces not conforming to specification in the lot = $1 - P = 0.05$

Sample proportion $\mathrm{p} = \frac{200 - 18}{200} = 0.91$

We take the null hypothesis that the proportion of pieces conforming to specification is 95% i.e., $H_0 : P = 0.95$

Alternative hypothesis $H_1 : P < 0.10$ (left-tailed test)

Since the sample size is large (≥ 30), therefore the Z test can be applied to test hypothesis.

Under the null hypothesis, the test statistic is given by

$$\mathrm{Z} = \frac{\mathrm{p} - \mathrm{P}}{\sqrt{\frac{\mathrm{PQ}}{\mathrm{n}}}}$$

$$= \frac{0.91 - 0.95}{\sqrt{\frac{0.95 \times 0.91}{200}}}$$

$$= \frac{-0.04}{0.015}$$

$$= -2.6$$

$$\therefore |Z| = 2.6$$

The critical value of $|Z|$ at $\alpha = 0.05$ for one-tailed test = 1.645.

Since the calculated value of $|Z|$ is greater than the critical value (i.e., 2.6 >1.645) therefore we reject the null hypothesis at 5% level and conclude that the claim of the manager is not valid.

[Alternatively, without taking the modulus we may compare the calculated value of Z with the critical value of $Z = -1.645$ at

$\alpha = 0.05$ for left-tailed test. Since $-2.6 < -1.645$, the null hypothesis is to be rejected and thus both types of comparison lead us to the same decision.]



CHECK YOUR PROGRESS

Q 4: State whether the following statements are true or false:

(i) A tentative assumption about a population parameter is called the null hypothesis.

(ii) The equality part of the expression (either $\geq, \leq,$ or $=$) always appears in the null hypothesis.

(iii) The estimated variance of sampling distribution of a statistic is called standard error.

(iv) The test statistic to test $\mu = \mu_0$ for large sample is z-test.

(v) The usual notation for the standard error of the sampling distribution is σ/n.

(vi) Suppose for a particular hypothesis $\alpha = 0.05$ and $\beta = 0.10$. Then the power of this hypothesis test is 0.90.

(vii) If 2.81 is the critical value of test-statistic z, then the significance level of the test is 0.05.

(viii) The large sample test for testing $P = P_0$ for normal population is t-test.

(viii) For test of hypothesis $H_0 : \mu = \mu_0$ against $H_1 : \mu > \mu_0$, the critical region at $\alpha = 0.05$ and $n \geq 30$, is $z > 1.96$.

(ix) When using the sample proportion $\bar{p}$ to test the hypothesis, the standard error of $\bar{p}$ is $\sqrt{\bar{p}\bar{q}/n}$.

(x) Sampling distribution will be approximately normal if sample size is small.

P – Value Approach for Hypothesis Testing: This approach is applied in hypothesis testing for large samples and sometimes referred to as the observed level of significance. This approach is very advantageous especially after the introduction of many statis-

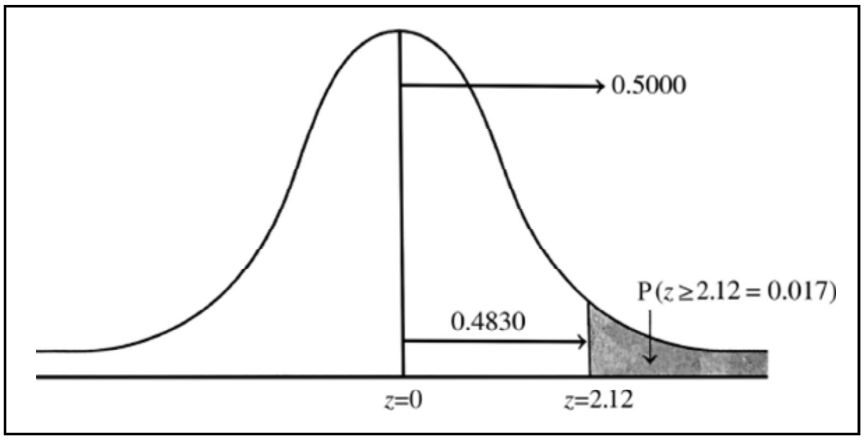
tical software packages. The $P-$ value defines the smallest value of α for which the null hypothesis can be rejected. The decision rule for accepting or rejecting a null hypothesis based on the $P-$ value as discussed below:

Reject the null hypothesis H_0 when the value is $<\alpha$, otherwise, accept the null hypothesis H_0

For example, suppose a researcher conducting a hypothesis test with rejection region on the right tail of the distribution obtains an observed test statistic value of 2.12. From the normal table, the corresponding probability area for z value 2.12 is 0.4830. Therefore, the probability of getting a value greater than or equal to 2.12 is 5.0000-0.4830=0.017 (as shown in the following figure)

In this case, the p value is 0.017. For α =0.05 and α=0.01, this value falls under the rejection region (α=0.05, 0.017< 0.05 and α=0.01, 0.017< 0.01), so the null hypothesis will be rejected at 0.05 0.01 levels of significance. At α=0.01, the researcher cannot reject the null hypothesis for the value of equal to 0.017 because α=0.01< 0.017.

Figure 6.6: Probability of p (z is greater than or equal to 2.12)



The procedure of calculating the p value for a two-tailed test is slightly different. In a two-tailed test, the rejection region falls in both the tails of the normal distribution. For example, at α=0.05, the rejection region is located in both the tail areas of the distribution in terms of 0.025% (0.025 +0.025=0.05) area in both the tails of the

distribution. The researcher compares this value to the computed value for accepting or rejecting the null hypothesis. For this comparison, instead of splitting the value, we double the value and then compare that value with the value. In the above mentioned example, the researcher conducted a hypothesis test with rejection region on both the tails of the distribution and obtained the observed test statistic as 2.12 and corresponding value as 0.017. So, instead of splitting value in two parts, the researcher should double the value, i.e., (0.017×2=0.034). So, for a two-tailed test, this value (0.034) is compared with the values 0.05 and 0.01. Hence, at α =0.05 and α =0.01 the null hypothesis will be rejected but at α =0.01 the null hypothesis will be accepted (0.01< 0.034).

Example 7: A marketing research firm conducted a survey 10 years ago and observed that the average household income of a particular geographic region is Rs. 10,000. Mr. Gupta, who has recently joined the firm as a vice president has expressed doubts about the accuracy of the data. For verifying the data, the firm has decided to take a random sample of 200 households that yield a sample mean (for household income) of Rs. 10,200. Assume that the population standard of the household income is Rs. 1200. Use the p -value approach to verify Mr. Gupta's doubts. (Use $\alpha = 0.01$ as the level of significance.)

Solution: We set up the null and alternative hypothesis given by

Null hypothesis $H_0 : \mu = 10{,}000$

Alternative hypothesis $H_1 : \mu \neq 10{,}000$

Since the sample size is large (≥ 30), therefore the Z test can be applied to test hypothesis.

Under the null hypothesis, the test statistic is given by

$$Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

$$= \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

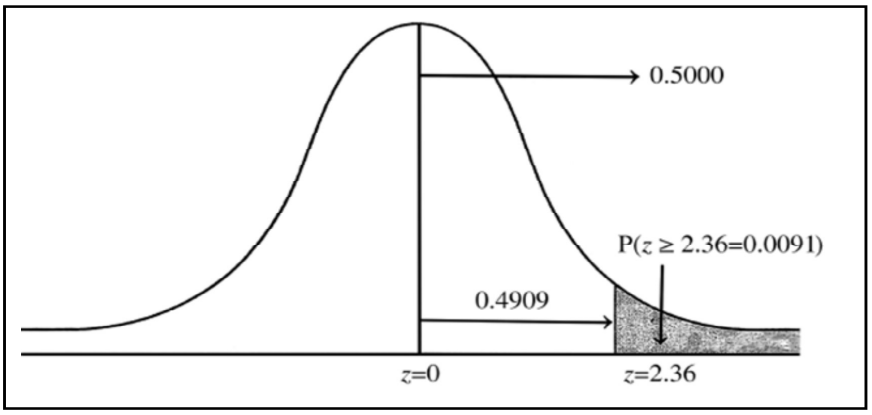
(since the population S.D. σ is unknown, therefore it can be replaced by the sample S.D. s)

$$= \frac{10{,}200 - 10{,}000}{\frac{1200}{\sqrt{200}}} = \frac{200}{84.85} = 2.36$$

Therefore the observed test statistic is computed as 2.36.

From the normal table, the corresponding probability area for Z value = 2.36 is 0.4909. So, the probability of obtaining a Z value greater than or equal to 2.36 is 0.5000-0.4909=0.0091 (as obtained from the following figure).

Figure 6.7: Probability of p (z is greater than or equal to 2.36)



For a two-tailed test, the obtained area value 0.0091 is to be multiplied by 2 and thus producing the value 0.0091×2=0.0182. Now since, $0.01 < 0.0182$, therefore we may accept the null hypothesis.

6.7.3 Hypothesis Testing for Difference between Two Population Means

In the real-world business environment, business researchers often encounter situations where they need to compare two means taken from two different samples drawn from two populations. On the basis of the samples researchers make an inference about the difference in the means of the two populations. For example, suppose a researcher is interested to conduct a survey to study the consumer

satisfaction for a particular brand of product in two different cities, Guwahati and Dibrugarh. In order to accomplish this, the researcher selects two different random samples from the two cities taken in the study. By using the procedures of testing of hypothesis, finally, the researcher draws a statistical conclusion regarding the consumer satisfaction in the two cities. Now, let $\overline{x_1}$ and $\overline{x_2}$ be the means of two large random samples of sizes n_1 and n_2 drawn from two populations with means μ_1 and μ_2 and standard deviations σ_1 and σ_2. We want to test the hypothesis whether the means of the two populations are equal or if there exists any significant difference between the population means.

Thus the null hypothesis is $H_0 : \mu_1 = \mu_2$

The alternative hypothesis is $H_1 : \mu_1 \neq \mu_2$

Assuming the null hypothesis to be true, the test statistic

$$Z = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

The decision rule is same as discussed in the earlier units.

Example 8: An investigation of the relative merits of two kinds of flashlight batteries showed that a random sample of 100 batteries of brand A lasted on the average for 35.5 hours with a standard deviation of 1.8 hours, while a random sample of 80 batteries of brand B lasted on the average 36.8 hours with a standard deviation of 1.5 hours. Do the two brands differ significantly in quality? Use $\alpha = 0.05.$

Solution: Let μ_1 and μ_2 denote the means of two populations.

Let us set up the null hypothesis that the two brands do not differ significantly in quality, i.e., $H_0 : \mu_1 = \mu_2$

and an alternative hypothesis $H_1 : \mu_1 \neq \mu_2$

Since the sample size is large (≥ 30), therefore the Z test can be applied to test hypothesis.

Under the null hypothesis, the test statistic is given by

$$Z = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

$$= \frac{36.5 - 36.8}{\sqrt{\frac{3.24}{100} + \frac{2.25}{80}}} = \frac{-0.3}{\sqrt{0.0324 + 0.0281}} = -1.2195$$

$$\therefore |Z| = 1.2195$$

The critical value of $|Z|$ for two-tailed test at $\alpha = 0.05$ is 1.96.

Since the calculated value of $|Z|$ is less than the critical value of $|Z|$ therefore we may accept the null hypothesis and conclude that there is no significant difference between the average life times of the two brands of batteries.

Example 9: The means of two large samples of size 1000 and 2000 are found to be 67.5 and 68.0 respectively. Test the equality of the two populations each with S.D. 2.5.

Solution: Let us set up null and alternative hypothesis as:

$H_0 : \mu_1 = \mu_2$

$H_1 : \mu_1 \neq \mu_2$

Where μ_1 is the mean of the first population and μ_2 is the mean of the second population.

Since the sample size is large (≥ 30), therefore the Z test can be applied to test hypothesis.

Under the null hypothesis, the test statistic is given by

$$Z = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

$$= \frac{67.5 - 68.0}{\sqrt{\frac{6.25}{1000} + \frac{6.25}{2000}}} = \frac{-0.5}{\sqrt{0.00625 + 0.00313}} = -5.16$$

$$\therefore |Z| = 5.16$$

Since the calculated value of $|Z|$ is greater than the critical value of $|Z|$ therefore we reject the null hypothesis and conclude that there is significant difference between population means.



CHECK YOUR PROGRESS

Q 5: A sample of 40 observations is selected from one population with S.D. (σ) 5. The sample mean is 102. A sample of 50 observations is selected from a second population with S.D. (σ) 6. The sample mean is 99. Conduct the following test of hypothesis (use $\alpha = 0.05$)

(a) Is this a one tailed or a two-tailed test? (b) State the decision rule. (c) Compute the value of the test statistic. (d) What is your decision regarding H_0?

6.7.4 Hypothesis Testing for Difference between Two Population Proportions

At times, we come across certain situations where we are interested to study whether or not there exist significant differences between the proportions of two groups of, say, consumer in respect of a certain activity. Let p_1 and p_2 be the proportions of a certain attribute in two large random samples of sizes n_1 and n_2 drawn from two populations having proportions P_1 and P_2. On the basis of the samples, we want to test whether the two population proportions are equal or if there exist any significant difference between the population proportions. Thus we will have null and alternative hypotheses as:

$$H_0 : P_1 = P_2$$
$$H_1 : P_1 \neq P_2$$

Assuming the null hypothesis to be true, the test statistic

$$Z = \frac{P_1 - P_2}{\sqrt{\frac{P_1 Q_1}{n_1} + \frac{P_2 Q_2}{n_2}}} \sim N(0,1)$$

where $Q_1 = 1 - P_1, \quad Q_2 = 1 - P_2$.

The procedure of taking decision is same as discussed earlier.

Note: (i) If P_1 and P_2 remains unknown, then they can be estimated by the corresponding sample proportions p_1 and p_2.

(ii) Sometimes, population proportion can be estimated by pooling the two sample proportions p_1 and p_2 into a single proportion p as given

$$p = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$$

In this case, the test statistic Z can be restated as

$$Z = \frac{p_1 - p_2}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}, \quad q = 1 - p$$

Example 10: In two populated states there are 30% and 25% respectively of fair haired people. Is this difference likely to be hidden in samples of 1200 and 900 respectively from the two populations?

Solution: Given $n_1 = 1200, \quad n_2 = 900$

$P_1 = 30\% = 0.30 \Rightarrow Q_1 = 0.70; \quad P_2 = 25\% = 0.25 \Rightarrow Q_2 = 0.75$

Now we set up the null hypothesis H_0 that the difference in population proportions is likely to be hidden in samples i.e.,

$$H_0 : P_1 = P_2$$

And the alternative hypothesis is

$$H_1 : P_1 \neq P_2$$

Since the samples are large, under H_0 the test statistic Z is given by

$$Z = \frac{P_1 - P_2}{\sqrt{\frac{P_1 Q_1}{n_1} + \frac{P_2 Q_2}{n_2}}} \sim N(0,1)$$

$$= \frac{0.30 - 0.25}{\sqrt{\frac{0.30 \times 0.70}{1200} + \frac{0.25 \times 0.75}{900}}}$$

$$= \frac{0.05}{\sqrt{0.000175 + 0.000208}} = \frac{0.05}{0.0196} = 2.55$$

Therefore the calculated value of Z =2.55

The critical value of Z at 5% level for two tailed test=1.96

Since the calculated value of Z > the critical value of Z , We reject the null hypothesis and hence, conclude that the difference in population proportions is not likely to be hidden in these samples.

Example 11: A Company has the head office at Calcutta and a branch at Guwahati. The personnel director wanted to assess if the workers at the two places would like the introduction of a new plan of work to enhance the quality of services and a survey was conducted for the purpose. Out of a sample of 500 workers at Calcutta 62% favoured the new plan. At Guwahati out of a sample of 400 workers 41% were against the new plan. Is there a significant difference between the two groups in their in their attitude towards the new $n_2 = 400, \quad p_2 = 59\% = 0.59 \Rightarrow q_2 = 1 - 0.59 = 0.41$ plan at 5% level?

Solution: Given

$n_1 = 500, \quad p_1 = 62\% = 0.62 \Rightarrow q_1 = 1 - 0.62 = 0.38$

Now we set up the null hypothesis that the population proportion of workers favouring the new plan in Calcutta and Guwahati is same, i.e. $H_0 : P_1 = P_2$,

The alternative hypothesis is

Since the samples are large, under the test statistic is given by

$$Z = \frac{p_1 - p_2}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}},$$

Where $p = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}, \quad q = 1 - p$

$$Z = \frac{0.62 - 0.59}{\sqrt{0.007 \times 0.393\left(\frac{1}{500} + \frac{1}{400}\right)}},$$

$$= \frac{0.03}{0.0327} = 0.917$$

Therefore the calculated value of Z =0.917

The critical value of Z at 5% level for two tailed test=1.96

Since the calculated value of $Z <$ the critical value of , We may accept the null hypothesis and hence, conclude that there is no significant difference between the two groups in their attitude towards the new plan.



6.8 LET US SUM UP

- Hypothesis testing is a well-defined procedure that helps us to decide objectively whether to accept or reject the hypothesis based on the information available from the sample. Testing of hypothesis involves five different steps.
- In case of testing of statistical hypothesis, there are two types of hypothesis tests namely two-tailed test and one-tailed test. Two-tailed tests contain the rejection region on both the tails of the sampling distribution of a test statistic. As different from a two-tailed test, a one-tailed test contains the rejection region on one tail of the sampling distribution of a test statistic.
- The testing of hypothesis involves two types of errors: Type I error and Type II error. A Type I error is committed by rejecting a null hypothesis when it is true. The possibility of committing Type I error is termed as level of significance and is denoted by α. On the other hand, a type II error is committed by accepting a null hypothesis when it is false. The likelihood of committing Type II error is denoted by β.
- Hypothesis testing can be performed by applying the approaches: the $z-$ value approach and $p-$ value approach. For testing hypothesis about a single population mean, z formula can be used if the sample

size is large $(n \geq 30)$ for any population with known population standard deviation σ. But in practical situations if σ is not available then it can be estimated by a sample standard deviation s. The p value defines the smallest value of for which the null hypothesis can be rejected.

- Hypothesis can also be performed for single population proportion as well as for the difference in two population proportions. In case of two random samples, on the basis of the difference in sample proportions, a researcher can estimate the difference in population proportions.



6.9 FURTHER READING

1) Gupta S.C. (1994). Fundamentals of statistics. New Delhi, Himalayan Publishing House.

2) Sharma, J.K. (2007). Business Statistics. New Delhi. Pearson Education Ltd.

3) Hazarika, P.L. (2016). Essential Statistics For Economics And Business Studies. New Delhi. Akansha Publishing House.

4) Lind, D.A., Marshal, W.G., Wathen, S.A. (2009) Statistical Techniques in Business and Economics. New Delhi. Tata McGraw Hill Education Private Limited.



6.10 ANSWERS TO CHECK YOUR PROGRESS

Ans to Q No 1: Hypothesis testing is the process of testing a statement or belief about a population parameter by the use of information collected from a sample (s).

Ans to Q No 2: Type I error is the probability of rejecting the null hypothesis when it is true. It is denoted by Greek letter α.

Ans to Q No 3: Type II error is the probability of accepting the null hypothesis when it is false. It is denoted by Greek letter β.

Ans to Q No 4: (i) T, (ii) T, (iii) F, (iv) T, (v) F, (vi) T, (vii) F, (viii) F, (ix) T, (x) F.

Ans to Q No 5: (a) Two-tailed test, (b) Reject H_0 $|z| > 1.96$,

(c) $z = \dfrac{102 - 99}{\sqrt{\dfrac{5^2}{40} + \dfrac{6^2}{50}}} = 2.59$ (d) Reject H_0



6.11 MODEL QUESTIONS

Q 1: Discuss the various steps involved in testing of hypothesis.

Q 2: What do you understand by null hypothesis, alternative hypothesis and level of significance?

Q 3: Define Type I error and Type II error.

Q 4: What is meant by one-tailed test and two-tailed test?

Q 5: The mean breaking strength of the cables supplied by a manufacturer is 1800 with a standard deviation of 100. By a new technique in the manufacturing process it is claimed that the breaking strength of the cables has increased. In order to test this claim a sample of 50 cables is tested. It is found that the mean breaking strength is 1850. Can we support the claim at 1% level of significance?

Q 6: The MacBurger restaurant chain claims that the mean waiting time of customers is 3 minutes with a population standard deviation of 1 minute. The quality- assurance department found in a sample of 50 customers at the Warren Road MacBurger that the mean waiting time was 2.75 minutes. At the 5% significance level, can we conclude that the mean waiting time is less than 3 minutes?

Q 7: A company manufacturing a certain type breakfast cereal claims that 60% of all housewives prefer that type to any other. A random sample of 300 housewives contains 165 who do prefer that type. At 5% level of significance, test the claim of the company.

Q 8: A man buys 50 electric bulbs of 'Philips' and 50 electric bulbs of 'HMT'. He finds that 'Philips' gave an average life of 1500 hours with a standard

deviation of 60 hours and 'HMT' bulbs gave an average life of 1512 hours with a standard deviation of 80 hours. Is there a significant difference in the mean life of the two makes of bulbes?

Q 9: 500 units from a factory are inspected and 12 are found to be defective, 800 units from another factory are inspected and 12 are found to be defective. Can it be concluded that at 5% level of significance production at the second factory is better than in first factory?

*** ***** ***

UNIT 7: TEST OF SIGNIFICANCE

UNIT STRUCTURE

7.1 Learning Objectives

7.2 Introduction

7.3 The Chi-Square (χ^2) Distribution

7.4 Applications of χ^2 Test

- 7.4.1 χ^2 Test of Goodness of Fit
- 7.4.2 χ^2 Test for Independence of Attributes
- 7.4.3 χ^2 Test for a Specified Population Variance
- 7.4.4 χ^2 Test of Homogeneity

7.5 Testing of Hypothesis in case of Small Samples (t-Test)

- 7.5.1 Characteristics of t-distribution
- 7.5.2 Applications of t-distribution
- 7.5.3 Hypothesis Testing for Single Population Mean
- 7.5.4 Hypothesis Testing for Difference between Two Population Means
- 7.5.5 Paired t-test for Difference between Two Population Means

7.6 Let Us Sum Up

7.7 Further Reading

7.8 Answer to Check Your Progress

7.9 Model Questions

7.1 LEARNING OBJECTIVES

After going tshrough this unit you will be able to

- introduce the Chi-square distribution for the comparison of more than two means or proportions
- use the Chi-square distribution to test whether two classifications of the same data are independent of each other
- use the Chi-square distribution for testing the hypothesis about a single population variance
- use t-test in small samples.

7.2 INTRODUCTION

In the previous unit we have learnt some statistical tests (using data from either one or two samples) that are used to test the hypothesis about the population parameters, with a number of business related examples. These tests are based on the assumption that the samples are drawn from a normally distributed population or, more appropriately, that the sample means are normally distributed. Such tests are known as parametric tests as the testing procedure involves the assumption about the type of population of parameters. On the other hand, in some cases, the researcher may not be sure of whether the population distribution is normal. The statistical tests that do not require prior knowledge about the population are termed as non-parametric tests. Chi-square test is one of the important non-parametric methods.

A non-parametric method, also known as distribution free test satisfies at least one of the following criteria:

(i) The method does not take into consideration any population parameter.

(ii) It is applied only on categorical data that are non-numerical and frequency counts of categories for one or more variables.

(iii) It does not depend on the form of the underlying population distribution, in particular, the requirement of normality.

In this unit we will focus on χ^2 (Chi-square) test. This test enables us to test whether more than two population proportions are equal. In fact, it allows to do a lot more than just test for the equality of several proportions. If we classify a population into several categories with respect to two attributes (such as age and job performance), we can then employ a chi-square test to test whether the two attributes are independent of each other.

7.3 THE CHI-SQUARE (χ^2) DISTRIBUTION

The symbol χ stands for the Greek letter 'chi'. This test was first developed by Karl Pearson in testing of hypothesis in 1900. The sampling distribution of χ^2 is called χ^2 distribution. This distribution has only one parameter called the degrees of freedom (df) and the shape of the curve

depends on the degrees of freedom. The distribution is skewed to the right. Being a square quantity, χ^2 distributi3on can never be negative. In fact, χ^2 is a continuous probability distribution with range 0 to ∞ (Figure 14.1)

The probability density function of distribution is given by

$$f\left(\chi^2\right) = C\left(\chi^2\right)^{\frac{\nu}{2}-1} e^{-\frac{\chi^2}{2}}$$

where ν is is the degrees of freedom, C is a constant depending upon the degrees of freedom, and $e = 2.71828.$

χ^2 test statistic can be defined as below:

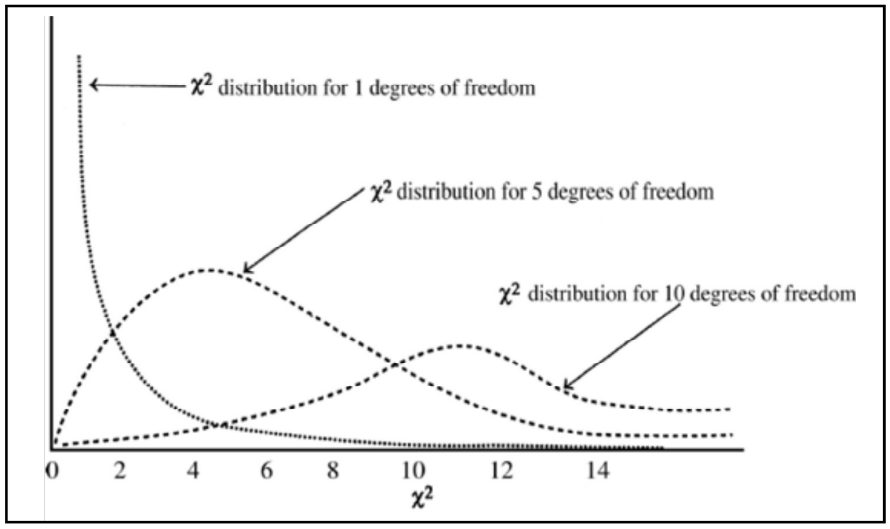
$$\chi^2 = \sum \frac{(O-E)^2}{E} \text{ with } df = k - 1 - c$$

where O is the observed frequency, E the expected frequency, k the number of categories, and c the number of parameters being estimated from the sample data.

The computed value of χ^2 is compared with the critical value of χ^2 at a specified level of significance and degrees of freedom. The decision to accept or reject the null hypothesis is given by

If $\chi^2_{cal} < \chi^2_{critical}$ then we may accept the null hypothesis, otherwise we reject the null hypothesis.

Figure 7.1: χ^2 Distribution with 1, 5 and 10 Degrees of Freedom



Properties of Chi-square test:

The following properties are useful while using χ^2 test statistic to analyze its sampling distribution:

(i) The density function of χ^2 does not involve any parameter and the shape of the curve depends upon the degrees of freedom only. The distribution becomes more and more symmetrical as the degrees of freedom increase and tends to normal distribution for high value of degrees of freedom ($\nu \geq 30$).

(ii) Mean = Degrees of freedom = ν, mode= $\chi^2 - 2$ and variance= 2ν

(iii) The values of χ^2 lie between 0 and ∞.

(iv) If $x_1, x_2, ..., x_n$ be a random sample drawn from a normal population with mean μ and variance σ^2, then

(a) $\sum_{i=1}^{n} (x_i - \mu)^2 / \sigma^2$ follows χ^2 -distribution with n degrees of freedom.

(b) $\sum_{i=1}^{n} (x_i - \bar{x})^2 / \sigma^2$ follows χ^2 -distribution with degrees of freedom $n-1$, $\bar{x}$ being the sample mean.

(v) For large degrees of freedom (ν), the standard variate

$$Z = \frac{\chi^2 - E(\chi^2)}{S.E.(\chi^2)} = \frac{\chi^2 - \nu}{\sqrt{2\nu}}$$

may approximately be considered as a standard normal variate.

Conditions for applications of Chi-square test:

The following conditions need to be satisfied before applying χ^2 as a test statistic for hypothesis testing:

(i) In a contingency table, an expected frequency of less than 5 in a cell is less than the frequency required to apply the test. In such cases we apply the χ^2 technique of 'pooling' which consists in adding the

may approximately be considered as a standard normal variate.

Conditions for applications of Chi-square test:

The following conditions need to be satisfied before applying χ^2 as a test statistic for hypothesis testing:

(i) In a contingency table, an expected frequency of less than 5 in a cell is less than the frequency required to apply the test. In such cases we apply the technique of 'pooling' which consists in adding the frequencies which are less than 5 with the preceding or succeeding frequency, so that the resulting sum is 5 or more and adjust for the degrees of freedom accordingly.

(ii) The sample should consist of at least 50 observations and should be drawn randomly from the population.

(iii) The data should be expressed in original units, rather than in percentage or ratio form.

(iv) The constraints on the cell frequencies, if any, should be linear such as $\sum O = \sum E = N$, the total frequency.

7.4 APPLICATIONS OF χ^2 TEST

The Chi-square test is not based on assumption of normality and can be used to both large and small samples. Chi-square distribution is a widely used since it can be employed in situations in which the normal distribution cannot be applied. Chi-square distribution has a number of applications, some of which are enumerated as below:

(i) to test goodness of fit,

(ii) to test the independence of attributes,

(iii) to test if the population has a specified value of standard deviation or variance.

(iv) to test if two or more populations are homogenous with respect to some characteristics of interest.

Critical Values of χ^2 :

From the χ^2 table (given in the Appendix) it may be observed that that the tabulated (critical) values of χ^2 increase as the degrees of freedom increases and the level of significance decreases.

Let us define the symbol $\chi_n^2(\alpha)$ as the value of χ^2 for $n\ d.f.$ such that the area to the right of this point is α, i.e.,

$$P\left[\chi^2 > \chi_n^2(\alpha)\right] = \alpha$$

is the rejection region. This is shown in the Figure 14.2.

Figure 7.2: Critical Value and Rejection region of χ^2 Statistic

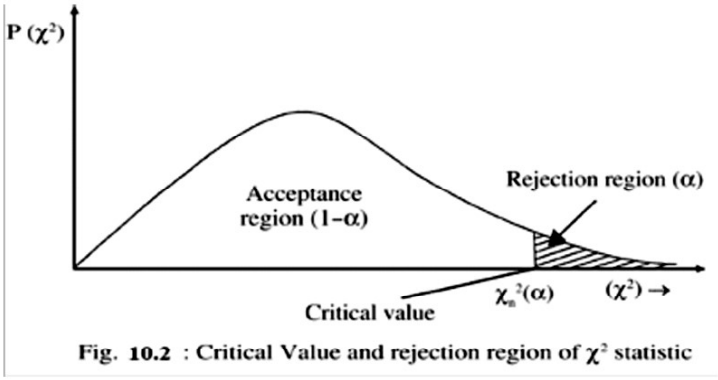


Fig. 10.2 : Critical Value and rejection region of χ^2 statistic



CHECK YOUR PROGRESS

Q 1. State whether the following statements are true or false:

(i) The critical value of a statistic is the value which cuts off the region for the rejection of null hypothesis.

(ii) The χ^2-test requires that the sampling distribution should be normally distributed.

(iii) If null hypothesis is accepted, then it is concluded that data analysis did not produce significant results.

(iv) In order to reject the null hypothesis $\chi^2_{cal} > \chi^2_{table}$

(v) The sampling distribution is a family of curves, the distribution of which varies with the degrees of freedom.

Now we discuss various applications of χ^2 :

7.4.1 χ^2 Test of Goodness of Fit

χ^2 test enables us to ascertain whether the known probability distributions such as Binomial, Poisson and Normal distributions fit with an observed sample observation. Thus χ^2 test for goodness of fit, a statistical test, provides a platform that can be

used to ascertain whether theoretical probability distributions coincide with empirical sample distributions. This test compares the theoretical (expected) frequencies with the observed (actual) frequencies to obtain the deviation between theoretical and observed frequencies. For the application of χ^2 test, first a theoretical distribution is hypothesized for a given population and then the test is carried out to make sure whether or not the sample data could come from the population of interest with the hypothesized theoretical distribution. The steps used for hypothesis testing can also be performed using the χ^2 goodness of fit test.

Steps to conduct χ^2 test for goodness of fit: The general steps to conduct a goodness of fit test for any hypothesized population distribution are summarized as below:

Step 1: Set up the null and alternative hypotheses as

Null hypothesis H_0 : No difference between the observed and expected sets of frequencies.

Alternative hypothesis H_1 : There is a difference.

Step 2: Calculate expected frequencies (E values) in each category by multiplying the category probability by the sample size ().

Step 3: Compute the test statistic

$$\chi^2 = \sum \frac{(\mathrm{O} - \mathrm{E})^2}{\mathrm{E}}$$

where O is the observed frequency, E the expected frequency.

Step 4: Using a specified level of significance α and $d.f. = n - 1$ provided that the number of expected frequencies are 5 or more for all categories, determine the critical value of χ^2 .

Step 5: Compare the computed value of χ^2 with the critical value of χ^2 at a specified level of significance and degrees of freedom. The decision to accept or reject the null hypothesis is given by

If χ^2_{cal} is less than the critical value $\chi^2_{\alpha,n-1}$ then we may accept the null hypothesis, otherwise we reject the null hypothesis.

Example 1: A company is concerned about the increasing violent altercations between its employees. The number of violent incidents

recorded by the management during six randomly selected months is given by

Months	Number of violent incidents
Jan	55
Feb	65
Mar	68
Apr	72
May	80
Jun	85

Test whether the data fits a uniform distribution. (Use $\alpha = 0.05$)

Solution: Let us set up null and alternative hypotheses as

Null hypothesis H_0 : Number of violent altercations are uniformly distributed over the months.

Alternative hypothesis H_1 : Number of violent altercations are not uniformly distributed over the months.

In order to test the null hypothesis, the appropriate test statistic is given by

$$\chi^2 = \sum \frac{(\mathrm{O} - \mathrm{E})^2}{\mathrm{E}}$$

where O is the observed frequency, E the expected frequency.

Now the expected frequencies will be calculated by dividing total observed frequencies by number of

months. Hence, the expected frequency $\frac{\sum \mathrm{O}}{6} = \frac{420}{6} = 70.$

Now to compute the test statistic we prepare the following table:

Months	O	E	$\frac{(O-E)^2}{E}$
Jan	55	70	3.2142
Feb	65	70	0.3571
Mar	68	70	0.0571
Apr	72	70	0.0571
May	80	70	0.9142
Jun	85	70	2.0571
Total	420		6.65

Thus $\chi^2 = 6.65$

$\therefore$ The calculated value of χ^2 is 6.65

The degrees of freedom= $n - 1 = 6 - 1 = 5$

The critical value of χ^2 at 5% level of significance for 5 degrees of freedom i.e., $\chi^2_{0.05,5} = 11.07$

Since the calculated value of is less than the critical value , therefore we may accept the null hypothesis and conclude that the number of violent altercations is uniformly distributed over the months.

Example 2: A brand manager is concerned that her brand's share may be unevenly distributed throughout the country. In a survey in which the country was divided into four geographical regions, a random sampling of 100 consumers in each region was surveyed, with the following information:

	Region				
	NE	NW	SE	SW	Total
Purchase the brand	40	55	45	50	190
Do not purchase	60	45	55	50	210
Total	100	100	100	100	400

Develop a table of observed and expected frequencies for this problem.

(i) Calculate the sample Chi-square value.

(ii) State the null and alternative hypothesis.

(iii) At $\alpha = 0.05$, test whether brand share is the same across the four regions.

Solution: To calculate the expected frequencies for the corresponding observed frequencies, we apply the following formula

(Row total × Column total)/Grand total.

The calculations of expected frequencies for the corresponding observed frequencies are given below:

		O	E	$O-E$	$(O-E)^2$	$(O-E)^2/E$
Row 1	NE	40	47.5	-7.5	56.25	1.18
	NW	55	47.5	7.5	56.25	1.18
	SE	45	47.5	-2.5	6.25	0.13
	SW	50	47.5	2.5	6.25	0.13
Row 2	NE	60	52.5	7.5	56.25	1.07
	NW	45	52.5	-7.5	56.25	1.07
	SE	55	52.5	2.5	6.25	0.12
	SW	50	52.5	-2.5	6.25	0.12
Total						5.00

(i) The sample Chi-square statistic is given by

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

where O is the observed frequency, E the expected frequency.

$\therefore$ From the above table we have,

$$\chi^2 = 5.00$$

(ii) We state the null and alternative hypotheses as:

Null hypothesis H_0 : The brand share is evenly distributed.

Alternative hypothesis H_1 : The brand share is not evenly distributed.

(iii) The calculated value of χ^2 is 5.00

The degrees of freedom = $(r-1)(c-1) = (2-1)(4-1) = 3$

The critical value of χ^2 at 5% level of significance for 3 degrees of freedom i.e., $\chi^2_{0.05,3} = 7.815$

Since the calculated value of is less than the critical value, therefore we may accept the null hypothesis and conclude that the brand share is evenly distributed in all the four regions of the country.

Example 3: A survey of 800 families with 4 children each revealed the following distribution:

Test whether the data are consistent with the hypothesis that the binomial law holds and the chance of a male birth is equal to that of a female birth.

Solution: Here we set up the null hypothesis that the data are consistent with the binomial law of equal probability for male and female births, i.e.,

$p = q = \frac{1}{2}$, (where p denotes the probability of a male birth.)

Using the binomial probability law, the expected frequency $f(x)$ of x male births is given by:

Number of births		Number of families
Male	Female	
0	4	32
1	3	178
2	2	290
3	1	236
4	0	64

$f(x) = N.p(x) = N \times {}^{n}C_x\, p^x q^{n-x}$, where

$N = 800,\ n = 4,\ x = 0, 1, 2, 3, 4.$

$$= 800 \times {}^{4}C_x \left(\frac{1}{2}\right)^4$$

$= 50 \times {}^{4}C_x$ .....(*)

Thus putting $x = 0, 1, 2, 3, 4$ successively in (*), we can determine the expected binomial frequencies as given below:

Number of Male births (x)	Expected frequency $f(x)$
0	$50\times{}^4C_0=50\times1=50$
1	$50\times{}^4C_1=50\times4=200$
2	$50\times{}^4C_2=50\times6=300$
3	$50\times{}^4C_3=50\times4=200$
4	$50\times{}^4C_4=50\times1=50$
Total	800

Let us now apply χ^2 test to examine the goodness of fit of the given data to the above binomial distribution. Calculations are given below:

	Observed frequency(O)	Expected frequency (E)	$O-E$	$(O-E)^2$	$\frac{(O-E)^2}{E}$
0	32	50	-18	324	6.48
1	178	200	-22	484	2.42
2	290	300	-10	100	0.33
3	236	200	36	1296	6.48
4	64	50	14	196	3.92
Total	800	800	0		19.63

Thus $\chi^2 = 19.63$

$\therefore$ The calculated value of χ^2 is 19.63

The degrees of freedom= $5-1=4$

The critical value of χ^2 at 5% level of significance for 4 degrees of freedom i.e., $\chi^2_{0.05,4} = 9.488$

Since the calculated value of χ^2 is greater than the critical value $\chi^2_{0.05,4}$, therefore we reject the null hypothesis and conclude that the hypothesis of equal male and female births is not valid.

Example 4: The figures given below are (a) the observed frequencies of a distribution, and (b) the frequencies of the normal distribution having the same mean, standard deviation and total frequency as in (a). Apply the χ^2 test of goodness of fit.

(a)	1	12	66	220	495	792	924	792	495	220	66	12	1
(b)	2	15	66	210	484	799	944	799	484	210	66	15	2

Solution: Since we are given the observed frequencies and expected frequencies of the normal distribution, therefore we apply the χ^2 test of goodness of fit.

Now we set up the null hypothesis that normal distribution is a good fit to the given frequency distribution, i.e., the given observed frequencies and the expected normal frequencies do not differ significantly.

In order to test the null hypothesis, the appropriate test statistic is given by

$$\chi^2 = \sum \frac{(\mathrm{O}-\mathrm{E})^2}{\mathrm{E}}$$

where O is the observed frequency, E the expected frequency.

Now to compute the χ^2 test we prepare the following table:

	O	E	$O-E$	$(O-E)^2$	$(O-E)^2/E$
	1, 12	2, 15	-4	16	0.94
	66	66	0	0	0.00
	220	210	10	100	0.48
	495	484	11	121	0.25
	792	799	-7	49	0.06
	924	944	-20	400	0.42
	792	799	-7	49	0.06
	495	484	11	121	0.25
	220	210	10	100	0.48
	66	66	0	0	0.00
	12, 1	15, 2	-4	16	0.94
Total	4096	4096	0		3.88

Thus $\chi^2 = 3.88$

$\therefore$ The calculated value of χ^2 is 3.88

The degrees of freedom=13-5=8

(The 13 original classes have been reduced to 11 by grouping, thus 2 degrees of freedom are lost in pooling. In addition, the mean, the standard deviation and total frequency of the original distribution

have been used in calculating the theoretical frequencies, thus introducing three restraints. The number of degrees of freedom is accordingly 8.)

The critical value of χ^2 at 5% level of significance for 8 degrees of freedom i.e., $\chi^2_{0.05,8} = 15.507$

Since the calculated value of χ^2 is much less than the critical value $\chi^2_{0.05,8}$, it is highly non-significant. Hence, we may accept the null hypothesis and conclude that the normal distribution is a good fit to the given distribution.



CHECK YOUR PROGRESS

Q 2: In a particular chi-square goodness-of-fit test there are four categories and 200 observations. Use the 5% significance level.

(a) How many degrees of freedom are there?

(b) What is the critical value of chi-square?

Q 3: The null hypothesis and alternative hypothesis are H_0 : The frequencies are equal and H_1 : The frequencies are not equal

Category	f
A	10
B	20
C	30

(a) State the decision rule ($\alpha = 0.05$), (b) Compute the value of Chi-square.

(c) What is your decision regarding

7.4.2 χ^2 Test for Independence of Attributes

In many business situations, a market researcher might be interested to study the relationship between two variables or to check whether they are independent of each other. For example, an edible oil company may be interested in knowing whether the purchase of

oil is independent of the customer's age or whether it is dependent on the customer's age. These are two different situations and consequently the company needs to frame a production and selling strategy. Similarly following are the situations where the Chi-square test of independence can be employed to analyse attributes with multiple categories.

- Whether university students classified by gender are independent of courses of study.
- Whether wage-earners classified by education level are independent of income.
- Whether a product produced is acceptable or not is independent of the shifts in which it was produced.

Thus when observations are classified on the basis of two qualitative variables or attributes and arranged in a table, the resulting table is known as a contingency table (Table 14.1). χ^2 test of independence uses this contingency table for determining independence of two qualitative variables.

Table 14.1: Contingency table

Variable B	Variable A				Total
	A_1	A_2	.....	A_k	
B_1	O_{11}	O_{12}	.....	O_{1k}	R_1
B_2	O_{21}	O_{22}		O_{2k}	R_2
.	.	.		.	.
.	.	.		.	.
.	.	.		.	.
B_r	O_{r1}	O_{r2}	.....	O_{rk}	R_r
Total	C_1	C_2		C_k	N

It is to be noted that the variables A and B have been classified into mutually exclusive categories. The value O_{ij} is the observed frequency for the cell in ith row and jth column. R_i is the total of the i^{th} row and C_j is the total of the j^{th} column. The row and column totals are added up to get a grand total (N), which represents the sample size. It is to be noted that to apply the Chi-square test we need to calculate the expected frequency.

The calculation of expected frequency, E_{ij} , corresponding to an observed frequency O_{ij} in row i and column j is based on the concept of multiplicative law of probability. Using this concept of probability the expected frequency, , is calculated as

$$E_{ij} = \frac{\text{i}^{\text{th}}\ \text{Row total}}{\text{sample size}} \times \frac{\text{j}^{\text{th}}\ \text{column total}}{\text{sample size}} \times \text{Grand total}$$

$$= \frac{\text{R}_\text{i}}{\text{N}} \times \frac{\text{C}_\text{j}}{\text{N}} \times \text{N} = \frac{\text{R}_\text{i} \times \text{C}_\text{j}}{\text{N}}$$

Thus expected frequency for any cell can be computed by using the formula

$$\text{E}_{\text{ij}} = \frac{\text{Row total} \times \text{Column total}}{\text{Grand total}}$$

The analysis of a two-way contingency table helps to determine whether the two variables are unrelated or independent of each other. Thus the null hypothesis for a Chi-square test of independence is that the two variables are independent.

Under the null hypothesis, the χ^2 test statistic is defined as

$$\chi^2 = \sum \frac{(\text{O} - \text{E})^2}{\text{E}}$$

The degrees of freedom for the Chi-square test of independence are given by

Degrees of freedom= (Number of rows-1) × (Number of columns-1)

$$= (r-1) \times (c-1)$$

Example 5: A certain drug is claimed to be effective in curing colds. In an experiment on 500 persons with cold, half of them were given the drug and half of them were given sugar pills. The patients' reactions to the treatment are recorded in the following table:

	Consequence			
Treatment	Helped	Reaction	No effect	Total
Drug	150	30	70	250
Sugar pills	130	40	80	250
Total	280	70	150	500

On the basis of the data, can it be concluded that there is a significant difference in the effect of the drug and sugar pills?

Solution: We state the null hypothesis H_0 that there is no significant difference in the effect of the drug and sugar pills. On the basis of this hypothesis expected frequencies are calculated as:

Treatment	Consequence Helped	Reaction	No effect	Total
Drug	$\frac{250\times280}{500}=140$	$\frac{250\times70}{500}=35$	250-(140+O5)=75	250
Sugar pills	140	35	75	250
Total	280	70	150	500

Now we arrange the observed and expected frequencies in the following table in order to calculate the χ^2 test statistic.

Observed (O)	Expected (E)	$(O-E)$	$(O-E)^2$	$(O-E)^2/E$
150	140	10	100	0.714
130	140	-10	100	0.714
30	35	-5	25	0.714
40	35	5	25	0.714
70	75	-5	25	0.333
80	75	5	25	0.333
Total				3.522

Thus the χ^2 test statistic is

$$\chi^2=\sum\frac{(O-E)^2}{E}=3.522$$

Number of degrees of freedom= (r-1) (c-1) = (2-1) (3-1) =2

The critical value of the χ^2 at 5% level of significance for 2 degrees of freedom = 5.99

Since the calculated value of the is less than the critical value, therefore we may accept the null hypothesis. Hence conclude that drug is no better than sugar pills in curing colds.

Example 6: Two researchers adopted different sampling techniques while investigating the same group of students to find the number of students falling in different intelligence levels. The results are as follows:

Researcher	Number of students in each level				Total
	Below average	Average	Above Average	Genius	
X	86	60	44	10	200
Y	40	33	25	2	100
Total	126	93	69	12	300

Would you say that the sampling techniques adopted by the two researchers are significantly different?

Solution: We state he null hypothesis that the data obtained are independent of the sampling techniques adopted by the two researchers.

Under the null hypothesis of independence, the expected frequencies are calculated as:

$$E_{11} = \frac{126 \times 200}{300} = 84; \qquad E_{12} = \frac{93 \times 200}{300} = 62; \qquad E_{13} = \frac{69 \times 200}{300} = 46$$

Now we arrange the observed and expected frequencies in the following table in order to calculate the χ^2 test statistic.

Observed (O)	Expected (E)	$(O-E)$	$(O-E)^2$	$(O-E)^2/E$
86	84	2	4	0.048
60	62	-2	4	0.064
44	46	-2	4	0.087
10	8	2	4	0.500
40	42	-2	4	0.095
33	31	2	4	0.129
25	23	2	4	0.174
2	4	-2	4	1.000
Total	300	0		2.097

Thus the χ^2 test statistic is

$$\chi^2 = \sum \frac{(O-E)^2}{E} = 2.097$$

Number of degrees of freedom= (r-1)(c-1)=(4-1)(2-1)=3

The critical value of the χ^2 at 5% level of significance for 3 degrees of freedom = 7.82

Since the calculated value of χ^2 the is less than the critical value, therefore we may accept the null hypothesis. Hence conclude that the sampling techniques adopted by the two investigators do not differ significantly.



CHECK YOUR PROGRESS

Q 4: A social scientist sampled 140 people and classified them according to income level and whether or not they played a state lottery in the last month. The sample information is reported below. Is it reasonable to conclude that playing the lottery is related to income level? Use

the 5% significance level.

	Income			
	Low	Middle	High	Total
Played	46	28	21	95
Did not play	14	12	19	45
Total	60	40	40	140

(a) What is this table called?

(b) State the null and alternative hypotheses.

(c) What is the degrees of freedom?

(d) What is the decision rule?

(e) Determine the value of χ^2.

(f) Make a decision about H_0.

7.4.3 χ^2 Test for a Specified Population Variance

Let $x_1, x_2, ..., x_n$ be a random sample of size n drawn from a normal population with mean μ and variance σ^2. We want to test the hypothesis that the population variance has a specified value.

We set up the null hypothesis and the alternative hypothesis as

$$H_0 : \sigma^2 = \sigma_0^2 \text{ and } H_1 : \sigma^2 > \sigma_0^2$$

Then under the null hypothesis, the test statistic

$$\chi^2 = \sum \frac{\left(x - \bar{x}\right)^2}{\sigma^2} = \frac{ns^2}{\sigma^2}$$

follows χ^2 distribution with $(n-1)\ d.f.$, where

$s^2 = \frac{\sum \left(x - \bar{x}\right)^2}{n}$, denotes the sample variance.

By comparing the calculated value of χ^2 with the tabulated value of for at the specified level of significance, we may accept or reject the null hypothesis at that level of significance.

Remarks: 1. test given by

$$\chi^2 = \frac{ns^2}{\sigma^2}$$

can be applied only if the population is normal.

2. If the sample size is large, say, greater than 30, then we can use Fisher's approximation, viz.,

$$\sqrt{2\chi^2} \sim N\left(\sqrt{2n-1}, 1\right)$$

And then apply the normal test by considering the statistic

$$Z = \sqrt{2\chi^2} - \sqrt{2n-1} \sim N(0,1)$$

Example 7: A battery manufacturer claims that the lives of his batteries have a standard deviation of 0.9 year. Can we accept manufacture's claim if the lives of a random sample of 10 batteries have a standard deviation of 1.2 years? (Use a 5% level of significance)

Solution: We set up the null and alternative hypotheses as

Null hypothesis $H_0 : \sigma = 0.9\ years$

Alternative hypothesis $H_1 : \sigma > 0.9\ years$

Under the null hypothesis, the test statistic

$\chi^2 = \frac{ns^2}{\sigma^2}$ ~Chi-square distribution with $(n-1)\ d.f.$

Now $\chi^2 = \frac{10 \times (1.2)^2}{(0.9)^2} = 17.78$

Degrees of freedom= n-1 =10-1=9

The critical value of χ^2 at 5% level for 9 d.f. i.e., $\chi^2_{0.05}(9) = 16.92.$

Since the critical value of χ^2 is greater than the tabulated value of value of χ^2 at 5% level for 9 d.f., we reject the null hypothesis and conclude that the standard deviation of life of the batteries is more than 0.9 years.

Example 8: The weights in kg. of 10 students are given below:

38, 40, 45, 53, 47, 43, 55, 48, 52, 49

Can we say that the variance of the distribution of weights of all students from which the above sample of 10 students was drawn, is equal to 20 square kg.?

Solution: We set up the null and alternative hypotheses as:

Null hypothesis $H_0 : \sigma^2 = 20\ sq.\ kg.$

Alternative hypothesis $H_1 : \sigma^2 > 20\ sq.\ kg.$

Under the null hypothesis, the test statistic

$\chi^2 = \frac{ns^2}{\sigma^2}$ ~Chi-square distribution with $(n-1)\ d.f.$

Now we prepare the following table to calculate the sample variance

Weight in kg. (x)	$(x-\bar{x})$ $x-47$	$(x-\bar{x})^2$
38	-9	81
40	-7	49
45	-2	4
53	6	36
47	0	0
43	-4	16
55	8	64
48	1	1
52	5	25
49	2	4
Total=470		280

$$\bar{x} = \frac{\sum x}{n} = \frac{470}{10} = 47 \qquad s^2 = \frac{\sum (x-\bar{x})^2}{n} = \frac{280}{10} = 28$$

Now, $\chi^2 = \frac{ns^2}{\sigma^2} = \frac{10 \times 28}{20} = 14$

Degrees of freedom= n-1=10-1=9

The critical value of χ^2 i.e., $\chi^2_{0.05}(9) = 16.92$

Since the calculated value of χ^2 is less than the critical value of $\chi^2_{0.05}(9)$, therefore we may accept our null hypothesis and conclude that the variance of the distribution of weights of all students in the population is 20 sq. kg.

7.4.4 χ^2 Test of Homogeneity

χ^2 test of homogeneity is used to determine whether two or more independent variables are drawn from the same population or from different populations. In other words, χ^2 test of homogeneity is used to know whether two or more populations are homogeneous with respect to some characteristics of interest. For example, a researcher may be interested in knowing whether the employees

from three departments, production, finance, and personnel, feel the same about the requirements of the top management in terms of hard work expected from employees. The amount of hard work can also be classified into three groups, namely, very hard, hard, and easy going. In this situation, we can state a null hypothesis that the opinion of all the groups is the same about the requirement of hard work. In other words, the null hypothesis assumes that the three classifications are homogeneous in terms of their opinion about the amount of hard work required by the top management.

It may be noted that, in both χ^2 test of homogeneity and χ^2 test of independence, one is concerned with the cross tabulation of the data and hypotheses testing procedure is also same. The differences that exist between the χ^2 test of homogeneity and χ^2 test of independence are: (i) In χ^2 test of independence, a researcher is interested in testing whether two attributes are independent. On the other hand in, χ^2 intest of homogeneity, a researcher investigates whether two or more populations are homogenous with respect to some characteristic of interest. (ii) in χ^2 test of homogeneity, two or more independent samples are drawn from each population whereas in case of χ^2 test of independence, a single sample is drawn from a population.

Example 9: A television company has launched a new product with some advanced features. The company wants to know the opinion of consumers about this product with respect to four characteristics: preferred brand with new features did not prefer brand with new features, preferred only a few new features, and indifferent. The company has divided consumers into three groups—executives/ officers; businessmen, and private consultants. It has taken a random sample of size 459 and obtained the following:

Consumers	Executives/Officers	Businessmen	Private consultants	Row total
Preferred brand with new features	35	25	40	100
Did not prefer brand with new features	30	45	34	109
Preferred only a few new features	45	50	25	120
Indifferent	25	55	50	130
Column total	135	175	149	459

Use χ^2 test of homogeneity and draw inference from the data.

Solution: Let us set up null and alternative hypotheses as

Null hypothesis H_0 : Opinion of all the groups is the same about the product with new features.

And alternative hypothesis H_1 : Opinion of all the groups is not the same about the product with new features.

In order to test the null hypothesis, the appropriate test statistic is given by:

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

where O is the observed frequency, E the expected frequency.

To calculate the expected frequencies for the corresponding observed frequencies, we apply the following formula

(Row total × Column total)/Grand total.

Now for the calculation of the expected frequencies against the corresponding observed frequencies and to compute the test statistic we prepare the following table:

Observed (O)	Expected (E)	$(O-E)$	$(O-E)^2/E$
35	29.41	5.59	1.0625
30	32.06	-2.06	0.1324
45	35.29	9.71	2.6717
25	38.24	-13.24	4.5841
25	38.13	-13.13	4.5213
45	41.56	3.44	0.2847
50	45.75	4.25	0.3948
55	49.56	5.44	0.5971
40	32.46	7.54	1.7514
34	35.38	-1.38	0.0538
25	38.95	-13.95	4.9962
50	42.20	7.8	1.4417
420			22.4917

Thus the χ^2 test statistic is

$$\chi^2 = \sum \frac{(O-E)^2}{E} = 22.4917$$

The number of degrees of freedom $= (r-1)(c-1) = (4-1)(3-1) = 6$

The critical value of the χ^2 at 5% level of significance for 6 degrees of freedom = 12.59

Since the calculated value of the χ^2 is greater than the critical value, therefore we reject the null hypothesis and conclude that the opinion of all the groups is not the same about the product with new features. Hence, the company has to consider different groups and their needs separately.

Precautions about using the Chi-square test: In order to apply Chi-square hypothesis test properly, one has to be extremely careful and keep in mind certain precautions. First, a sample should be large enough. If the expected frequencies are too small, the value of Chi-square gets over estimated. This will result in the rejection of the null hypothesis in several situations. To overcome this problem, we must ensure that the expected frequency in any cell of the contingency table should not be less than 5. In case the expected frequency less than 5 in more than one cell, we can combine these to obtain the expected frequency of at least 5. Another point to note is that the calculations must be made with actual numbers in each cell and not with proportions or percentages. If the proportions or percentages were used, then the theoretical distribution would not be applicable.

When the calculated value of Chi-square turns out to be more than the critical value at a specified level of significance, we reject our null hypothesis. On the other hand, when Chi-square value is less than the critical value, we may accept our null hypothesis. However, when the Chi-square value turns out to be zero, we have to be extremely careful to confirm that there is no difference between the observed and the expected frequencies. Such a situation may sometimes arise on account of a faulty manner used in the data

collection.

In most of the cases, the problems of Chi-square involve simple calculations. However, for large sets of data the Chi-square test involves very comprehensive calculations. In all such situations, computer should be used. Several computer Statistics packages contain routines for carrying out Chi-square tests.



CHECK YOUR PROGRESS

Q 5: When chi-square test of homogeneity is useful?

.....

.....

.....

.....

.....

7.5 TESTING OF HYPOTHESIS IN CASE OF SMALL SAMPLES (t-Test)

We have already discussed that in case of large samples (i.e. $n \geq 30$,) the sampling distribution of the statistic is approximately normal. But in many decision making situations, due to time, money and other constraints, sometimes a researcher may have to take small sample of size less than 30 (i.e. $n < 30$,). In case of small samples also, the researcher can apply standard normal test for a specified population mean or equality of two population means provided the sample is drawn from a normal population where S.D. σ is known. But if σ remains unknown then standard normal test fails and in all such cases the sampling distribution turns out to be the student's 't' distribution. Sir William Gosset of Ireland in early 1900, under his pen name 'Student', developed a method for testing of hypothesis known as 't-test'. It is to be noted that the concept of degrees of freedom is associated with t-test. The graph of the 't' distribution approaches the normal curve as the degrees of freedom increase.

If $x_1, x_2, \ldots\ldots\ldots\ldots, x_n$ be a random sample of size drawn from a normal population with mean μ and variance σ^2, then the Student's 't' statistic is defined by

$$t = \frac{\bar{x} - \mu}{S/\sqrt{n}},$$

where $\bar{x}$ = sample mean

and $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$, sample mean square

is an unbiased estimator of σ^2. If s^2 be the sample variance given by

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

Then $ns^2 = (n-1)S^2$

Thus we have, $\frac{S}{\sqrt{n}} = \frac{s}{\sqrt{n-1}}$

$$\therefore \quad t = \frac{\bar{x} - \mu}{S/\sqrt{n}} = \frac{\bar{x} - \mu}{s/\sqrt{n-1}}$$

Here, 't' follows t-distribution with (n-1) d.f.

7.5.1 Characteristics of t-Distribution

The characteristics of t distribution are given below:

(1) It is an exact distribution.

(2) The distribution is unimodal with

Mean = Median = Mode = 0

(3) The t-distribution is less peaked than normal distribution at the centre and higher at both the ends.

(4) The shape of the -distribution depends on depends on the sample size. As increases, it approaches the normal distribution.

(5) The limits of t is $-\infty < t < \infty$.

(6) The critical values of at level of significance vary with the degrees of freedom.

7.5.2 Applications of t-Distribution

The -distribution has a number of applications in Statistics some of which are given below

(i) It is used to test the significance of single mean, population standard deviation being unknown.

(ii) It is used to test the significant difference between two population means under the assumption that population standard deviations being equal but unknown.

(iii) It is used to test the significance of an observed sample correlation coefficient.

Assumptions for Student's -test: The following assumptions are made in Student's -test:

(i) The parent population from which the sample is drawn is normal.

(ii) The drawn sample is random.

(iii) The population standard deviation σ remains unknown.

7.5.3 Hypothesis Testing for Single Population Mean

Suppose a random sample of size $n(n<30)$ with mean $\bar{x}$ is drawn from a population μ. We want to test the hypothesis that the population has the specified value μ_0 when the population S.D. σ is unknown.

Here we set up the null hypothesis $H_0 : \mu = \mu_0$.

Now depending upon the situation, against this hypothesis we may have any one of the following three alternative hypotheses

(i) $H_1 : \mu \neq \mu_0$ (Two-tailed test)

(ii) $H_1 : \mu > \mu_0$ (Right-tailed test)

(iii) $H_1 : \mu < \mu_0$ (Left-tailed test)

Since the sample is small, therefore under the null hypothesis the test statistic to be constructed is

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n-1}}}$$ follows t distribution with $\nu = (n-1)$ d f.

where $s =$ sample S.D., $\mu = \mu_0$, σ is the population standard deviation, n is the sample size and $\bar{x}$ is the sample mean.

Note: For large sample ($n \geq 30$), we have

S.E. ($\bar{x}$) = $\frac{s}{\sqrt{n}}$

On the other hand, for small sample ($n < 30$),

S.E. ($\bar{x}$) = $\frac{s}{\sqrt{n-1}}$

The critical values (i.e., tabulated values) of t distribution given in the table for 't' in the appendix are the critical values of for right-tailed test. The corresponding values for left-tailed test will just be the negative of these values. If the values of for a two-tailed test at level are to be determined then instead of α, we are to take $\alpha/2$.

If the alternative hypothesis is two-tailed as defined in (1), then the null hypothesis is to be rejected if the calculated value of $|t| > t_{\alpha/2}$. Again if the alternative hypothesis is one-tailed as defined in (2) or (3), the null hypothesis is to be rejected if the calculated value of $|t| > t_{\alpha}$.

Example 10: Royal Tyres has launched a new brand of tyres for tractors and claims that under normal circumstances the average life of the Tyres is 40,000 km. A retailer wants to test this claim and has taken a random sample of 8 Tyres that yields mean of 39,750 with a standard deviation of 2618.61. He tests the life of the Tyres under normal circumstance. (Use $\alpha = 0.05$)

Solution: Here the size of the sample $(i.e., n)$ is 8 (< 30) and the population standard deviation is unknown, so test can be applied to test the hypothesis.

Given $\bar{x} = 39750$ and $s = 2618.61$

Now, we set up the null and alternative hypotheses as:

Null hypothesis $H_0 : \mu = 40{,}000$

Alternative hypothesis $H_1 : \mu \neq 40{,}000$

Under the null hypothesis the test statistic is

$$t = \frac{x - \mu}{\frac{s}{\sqrt{n-1}}}$$

$$= \frac{39750 - 40000}{\frac{2618.61}{\sqrt{8-1}}}$$

$$= -\frac{250}{1070.677132} = -0.233$$

$$\therefore \quad |t| = |-0.233| = 0.233$$

Degrees of freedom= $n - 1 = 7$

The critical value of $|t_{0.05}(7)|$ for two-tailed test = 2.365

Since the critical value of $|t|$ is less than the critical value of $|t|$ at 5% level for 7 d.f., therefore we may accept the null hypothesis at 5% level of significance and conclude that the company's claim is valid under normal conditions.

Example 11: The mean price of shares of a company during the year 1992 was Rs. 85. In 1993, the mean price of shares of 20 randomly selected days is found to be Rs. 75 with a standard deviation of Rs. 10. Is there a significant difference in the prices of shares for the two years? (Given $|t_{0.05}(19)| = 2.093$)

Solution: Here the size of the sample $(i.e., n)$ is 20 (< 30) and the population standard deviation σ is unknown, so t test can be applied to test the hypothesis.

Given $\bar{x} = RS.75$ and $s = Rs.10$

Now, we set up the null and alternative hypotheses as:

Null hypothesis $H_0 : \mu = Rs.85$

Alternative hypothesis $H_0 : \mu \neq Rs.85$

Under the null hypothesis the test statistic is

$$t = \frac{x - \mu}{\frac{s}{\sqrt{n-1}}}$$

Which follows t distribution with $(n-1)$ d f.

Now $$t = \frac{75 - 85}{\frac{10}{\sqrt{20-1}}}$$

$$= \frac{-10}{\frac{10}{\sqrt{19}}} = -\sqrt{19} = -4.36$$

$$\therefore \quad |t| = |-4.36| = 4.36$$

Degrees of freedom = $20 - 1 = 19$

The critical value of $|t_{0.05}(19)|$ for two-tailed test = 2.093

Since the critical value of $|t|$ is greater than the critical value of $|t|$ at 5% level for 19 d.f., therefore we reject the null hypothesis at 5% level of significance and conclude that there is significant difference between the mean prices of shares in the two years.

Example 12: A fertilizer mixing machine is set to give 12 kg. of nitrate for every 100 kg. of fertilizer. Ten bags of 100 kg. each are examined. The percentage of nitrate so obtained is: 11, 14, 13, 12, 13, 12, 13, 14,11 and 12. Is there reason to believe that the machine is defective? (Use $\alpha = 0.05$)

Solution: Given $n = 10$

Let us consider the null and alternative hypotheses as

$H_0 : \mu = 12$ and $H_1 : \mu \neq 12$

We now calculate mean and standard deviation of the given sample as shown below:

Calculations of sample mean and standard deviation:

Variable x	$x-\bar{x}$ $=x-12.5$	$(x-\bar{x})^2$
11	-1.5	2.25
14	1.5	2.25
13	0.5	.25
12	-0.5	.25
13	0.5	.25
12	-0.5	.25
13	0.5	.25
14	1.5	2.25
11	-1.5	2.25
12	-0.5	.25
125	0	10.5

$$\bar{x}=\frac{\sum x}{n}=\frac{125}{10}=12.5 \text{ and } s=\sqrt{\frac{1}{n}\sum(x-\bar{x})^2}=\sqrt{\frac{10.5}{10}}=\sqrt{1.05}=1.025$$

Under null hypothesis the test statistic is

$$t=\frac{\bar{x}-\mu}{\frac{s}{\sqrt{n-1}}}$$

Which follows t distribution with $(n-1)$ d.f

Now $$t=\frac{12.5-12}{\frac{1.025}{\sqrt{10-1}}}=\frac{0.5}{\frac{1.025}{\sqrt{9}}}=\frac{0.5}{0.3417}=1.46$$

$\therefore \quad |t|=|1.46|=1.46$

Degrees of freedom= $10-1=9$

The critical value of $|t_{0.05}(19)|$ for two-tailed test = 2.262

Since the critical value of $|t|$ is less than the critical value of $|t|$ at 5%

level for 9 d.f., therefore we may accept the null hypothesis at 5% level of significance and conclude that the manufacturer's claim is valid, i.e., the machine is not defective.

7.5.4 Hypothesis Testing for Difference between Two Population Means

In case of two populations, our concern is to test if two independent samples have been drawn from two normal populations having the same means, the population standard deviations being equal. Let $\overline{x_1}$ and $\overline{x_2}$ be the means of two independent random samples of sizes n_1 and n_2 ($n_1 < 30, n_2 < 30$) drawn from two normal populations having means μ_1 and μ_2 respectively with equal but unknown standard deviations σ_1 and σ_2.

Thus the null hypothesis is $H_0 : \mu_1 = \mu_2$, i.e., samples have been drawn from the normal populations with same means.

The alternative hypothesis H_1 may be any one of the following depending upon the situation.

(1) $H_1 : \mu_1 \neq \mu_2$ (two-tailed)

(2) $H_1 : \mu_1 < \mu_2$ (left-tailed)

(3) $.H_1 : \mu_1 > \mu_2$ (right-tailed)

Assuming the null hypothesis to be true, the test statistic t is given by

$$t = \frac{\overline{x_1} - \overline{x_2}}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1+n_2-2}$$

i.e., the test statistic t follows t distribution with n_1+n_2-2 d.f

where $\overline{x_1} = \frac{1}{n}\sum x_1, \qquad \overline{x_2} = \frac{1}{n}\sum x_2$

and $S^2 = \frac{1}{n_1 + n_2 - 2}\left[\sum (x_1 - \overline{x_1})^2 + \sum (x_2 - \overline{x_2})^2\right]$

$= \frac{1}{n_1 + n_2 - 2}\left[n_1 s_1^2 + n_2 s_2^2\right]$; where $s_1^2 = \frac{1}{n_1}\sum (x_1 - \overline{x_1})^2$ and $s_2^2 = \frac{1}{n_2}\sum (x_2 - \overline{x_2})^2$

is an unbiased estimate of the common population variance σ^2 based on both the samples.

The decision to reject or accept the null hypothesis is similar to that used in one-sample tests.

Example 13: The IQs (intelligence quotients) of 16 students from one area of a city showed a mean of 107 with a standard deviation of 10, while the IQs of 14 students from another area of the city showed a mean of 112 with a standard deviation of 8. Is there a significant difference between the IQs of the two groups at 5% level of significance?

Solution:

Given $n_1 = 16, \quad n_2 = 14, \quad \overline{x_1} = 107, \quad \overline{x_2} = 112, \quad s_1 = 10$ and $s_2 = 8$

We set up null and alternative hypothesis as

Null hypothesis is $H_0 : \mu_1 = \mu_2$

Alternative hypothesis is $H_0 : \mu_1 \neq \mu_2$

Since sizes of both the samples are small and population standard deviations are unknown therefore to test the hypothesis we apply t test given by

$$t = \frac{\overline{x_1} - \overline{x_2}}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1+n_2-2}$$

i.e., the test statistic t follows t distribution with n_1+n_2-2 d.f.

Now

$$S^2 = \frac{1}{n_1 + n_2 - 2}\left[n_1 s_1^2 + n_2 s_2^2\right]$$

$$= \frac{1}{16+14-2}\left[16(10)^2 + 14(8)^2\right]$$

$$= 89.142857$$

Thus $S = 9.44$

$$\therefore \quad t = \frac{112-107}{9.44\sqrt{\frac{1}{16}+\frac{1}{14}}} = 1.45$$

The degrees of freedom = $n_1+n_2-2=28$

The critical value of $|t_{0.05}(28)|$ for two-tailed test = 2.02

Since the critical value of $|t|$ is less than the critical value of $|t|$ at 5% level for 28 d.f., therefore we may accept the null hypothesis at 5% level of significance and conclude that the difference between the two population means is not significant.

Example 14: In a test given to two groups of students, the marks obtained are as given below:

First group : 25 32 30 32 24 14 32

Second group: 24 34 22 30 42 31 40 30 32 35

Examine the significance of the difference between the arithmetic mean of the marks secured by the students of the above two groups.

Solution: Here we set up null and alternative hypotheses as

Null hypothesis is $H_0 : \mu_1 = \mu_2$ i.e., there is no significant difference between mean marks secured by students of the two groups.

Alternative hypothesis is $H_0 : \mu_1 \neq \mu_2$

Since sizes of both the samples are small and population standard deviations are unknown therefore to test the hypothesis we apply t test given by

$$t = \frac{\overline{x_1} - \overline{x_2}}{S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t_{n_1+n_2-2}$$

i.e., the test statistic t follows t distribution with n_1+n_2-2 d.f.

Now

$$S^2 = \frac{1}{n_1+n_2-2}\left[n_1 s_1^2 + n_2 s_2^2\right]$$

Calculations of $\overline{x_1}$, $\overline{x_2}$ and pooled sample standard deviations

are given below:

Calculation of $\overline{x_1}$, $\overline{x_2}$ and S

x_1	$x_1-\overline{x_1}$ $=x_1-27$	$(x_1-\overline{x_1})^2$	x_2	$x_2-\overline{x_2}$ $=x_2-32$	$(x_2-\overline{x_2})^2$
25	-2	4	24	-8	64
32	5	25	34	2	4
30	3	9	22	-10	100
32	5	25	30	2	4
24	-3	9	42	10	100
14	-13	169	31	-1	1
32	5	25	40	8	64
			30	-2	4
			32	0	0
			35	3	9
189		266	320		350

Here $\overline{x}_1=\frac{\sum x_1}{n_1}=\frac{189}{7}=27$ $\overline{x}_2=\frac{\sum x_2}{n_2}=\frac{320}{10}=32$

Sample variance, $s_1^2=\frac{1}{n_1}\sum(x_1-\overline{x_1})^2=\frac{266}{7}=38$

Sample variance, $s_2^2=\frac{1}{n_2}\sum(x_2-\overline{x_2})^2=\frac{350}{10}=35$

Now, $S=\sqrt{\frac{7\times38+10\times35}{7+10-2}}=6.41$

$$\therefore \quad t=\frac{27-32}{6.41\sqrt{\frac{1}{7}+\frac{1}{10}}}$$

$$=\frac{-5}{6.41\sqrt{\frac{17}{70}}}=-\frac{5}{6.41\times0.49}=-1.592$$

$$\therefore \quad |t|=1.592$$

The degrees of freedom = $n_1+n_2-2=15$

The critical value of $|t_{0.05}(15)|$ for two-tailed test = 2.131

Since the critical value of $|t|$ is less than the critical value of $|t|$ at 5% level for 28 d.f., therefore we may accept the null hypothesis at 5% level of significance and conclude that the mean marks obtained by the students of two groups do not differ significantly.

Example 15: Two salesmen, A and B, are employed by a company. To study if there is any significant difference between the average sales of the two salesmen, the HRD manager of the company conducted a survey yielding the following information:

	Salesman A	Salesman B
Number of sales	20	22
Average weekly sales (Rs. lakh)	30	25
Standard deviation (Rs lakh)	10	7

Solution: First we set up the null hypothesis that there is no significant difference between the average sales of the two salesmen. That is, $H_0 : \mu_1 = \mu_2$

The alternative hypothesis is $H_1 : \mu_1 \neq \mu_2$

Since sizes of both the samples are small and population standard deviations are unknown therefore to test the hypothesis we apply t test given by

$$t = \frac{\overline{x_1} - \overline{x_2}}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1+n_2-2}$$

i.e., the test statistic t follows t distribution with n_1+n_2-2 d.f.

Now

$$S^2 = \frac{1}{n_1 + n_2 - 2}\left[n_1 s_1^2 + n_2 s_2^2\right]$$

$$= \frac{1}{20 + 22 - 2}\left[20(10)^2 + 22(7)^2\right] = 76.95$$

Thus $S = 8.77$

$$\therefore \quad t = \frac{30 - 25}{8.77\sqrt{\frac{1}{20} + \frac{1}{22}}} = 1.84$$

The degrees of freedom = $n_1+n_2-2=40$

The critical value of $|t_{0.05}(40)|$ for two-tailed test = 2.021

Since the calculated value of $|t|$ is less than the critical value of $|t|$ at 5% level for 40 d.f., therefore we may accept the null hypothesis at 5% level of significance and conclude that the average sales by the two salesmen are not significantly different.



CHECK YOUR PROGRESS

Q 6: To the following information (a) State null and alternative hypotheses, (b) State the decision rule, (c) Compute the pooled estimate of the population variance, (d) Compute the test statistic. (e) What is your decision regarding H_0 ?

A random sample of 10 observations from one population revealed a sample mean of 23 and a S.D. of 4. A random sample of 8 observations from another population revealed a sample mean of 26 and a S.D. of 5. Use 5% level of significance to test if there is any significant difference between population means?

7.5.5 Paired t – Test for Difference between Two Population Means

In the t – test used for difference of means, the two samples were independent of each other whereas in case of paired t – test we have

(i) The sample sizes (n_1 and n_2) are equal.

(ii) The sample observations $(x_1, x_2, \ldots\ldots\ldots\ldots, x_n)$ and $(y_1, y_2, \ldots\ldots\ldots\ldots, y_n)$ are not completely independent but they are dependent in pairs i.e., the pairs $(x_1, y_1), (x_2, y_2), \ldots\ldots\ldots (x_n, y_n)$ correspond to the 1st, 2nd, ……,nth unit respectively.

For example, the management of a company plagued by poor productivity realizes the need to provide technical training to employees. It hires a researcher to measure the productivity levels of a sample of 25 employees. The productivity levels are measured again after a one-month technical training programme. In this kind of pre-and post-training study, which are selected before and after the study cannot be treated as independent because each observation in sample 1 is related to the observation in sample 2. The productivity scores obtained before training is related to the scores obtained after training because the two measurements are apply to the same person.

Now we explain the paired $t-$ test for difference between the means of two related populations as follows:

Suppose a business concern is interested to know if advertisement is really effective in promoting sales of a particular product. Let ($x_1, x_{2,}.....,x_n$) be a sample of sales of the product in departmental stores for a certain period before advertisement campaign and ($y_1, y_2,.....,y_n$) be the sample of corresponding sales of the same product and for period in the same departmental stores respectively. Now $(x_i, y_i), i = 1,2,.....,n$ is the pair of sales in the i[th] departmental store before and after advertisement. Suppose we want to test the significance of the difference between the sales, and.

Let $d_i = x_i - y_i, (i = 1,2,.....,n)$

We set up the null hypothesis that there is no significant change in sales after advertisement campaign. i.e., $H_0 : \mu_1 = \mu_2$.

The alternative hypothesis is $H_1 : \mu_1 \neq \mu_2$.

Under the null hypothesis, the test statistic is

$$t = \frac{\overline{d}}{S/\sqrt{n}} \sim t_{n-1}$$

where $d = x - y, \quad \bar{d} = \frac{1}{n}\sum d$

and $S^2 = \frac{1}{n-1}\sum\left(d - \bar{d}\right)^2 = \frac{1}{n-1}\left[\sum d^2 - \frac{\left(\sum d\right)^2}{n}\right]$

Example 16: An electronic company arranged a special training programme for one segment of its employees. In order to measure the change in the attitude of its employees after the training, the company has prepared a well-designed questionnaire consisting of 10 questions on a 1 to 5 rating scale (1 is strongly disagree and 5 is strongly agree). The company selected a random sample of 10 employees. The scores are given below:

Employees	Scores before training	Scores after training
1	25	32
2	26	30
3	28	32
4	22	34
5	20	32
6	30	28
7	22	25
8	20	30
9	21	25
10	24	28

Test whether there is a significant change in the attitude of employees after the training programme. (Use $\alpha = 0.10$)

Solution: Here, scores before training (x) and after training (y) are not independent but are paired together. Hence, to test the difference between the means of two related populations we apply the paired $t-$test.

We set up the null and alternative hypotheses as:

The null hypothesis $H_0 : \mu_1 = \mu_2$ i.e., there is no significant effect of the training.

The alternative hypothesis $H_1 : \mu_1 \neq \mu_2$.

Under the null hypothesis, the test statistic is

$$t = \frac{\bar{d}}{S/\sqrt{n}} \sim t_{n-1}$$

where $d = x - y, \quad \bar{d} = \frac{1}{n}\sum d$

and $$S^2 = \frac{1}{n-1}\sum\left(d - \bar{d}\right)^2 = \frac{1}{n-1}\left[\sum d^2 - \frac{\left(\sum d\right)^2}{n}\right]$$

Calculations of $\bar{d}$ and S^2 are given below:

Employees	Scores before training(x)	Scores after training(y)	$d = x - y$	d^2
1	25	32	-7	49
2	26	30	-4	16
3	28	32	-4	16
4	22	34	-12	144
5	20	32	-12	144
6	30	28	2	4
7	22	25	-3	9
8	20	30	-10	100
9	21	25	-4	16
10	24	28	-4	16
Total			-58	514

$$\therefore d = \frac{\sum d}{n} = \frac{-58}{10} = -5.8$$

$$S^2 = \frac{1}{10-1}\left[514 - \frac{(-58)^2}{10}\right] = \frac{1}{9}(514 - 336.4) = 19.733333$$

Thus $S = \sqrt{19.733333} = 4.44$

$$\therefore t = \frac{-5.8}{4.44/\sqrt{10}} = \frac{-5.8}{1.4047} = -4.13$$

$$|t| = 4.13$$

The degrees of freedom = 9

The critical value of $|t_{0.10}(9)|$ for two-tailed test = 1.833

Since the calculated value of $|t|$ is greater than the critical value of $|t|$ at 10% level for 9 d.f., therefore we reject the null hypothesis at 10% level of significance and conclude that there is a significant difference in the attitude of employees before and after the training.



CHECK YOUR PROGRESS

Q 7: To the following (a) State null hypothesis, (b) State the decision rule, (c) Compute the test statistic. (e) What is your decision regarding H_0?

An IQ test was administered to 5 persons before and after they were trained. The results are given below:

Candidate:	I	II	III	IV	V
IQ before training:	110	120	123	132	125
IQ after training:	120	118	125	136	121

Test if there is any change in IQ level after the training programme.

.....

.....

.....

.....

.....

.....



7.6 LET US SUM UP

- Statistical tests that do not require any prior information about the population are termed as non-parametric tests. This unit focuses on only χ^2 (Chi-square) distribution and the related χ^2 test. χ^2 distribution is the family of curves with each distribution being defined by the degree of freedom associated to it.
- χ^2 test can be used for a variety of purposes. χ^2 test provides a

platform that can be used to ascertain whether the theoretical probability distribution coincides with the empirical sample distribution. This is commonly known as χ^2 goodness of fit test. It can also be used to test the independence of independence of two variables. χ^2 test of independence makes use of the contingency table to determine the independence of two variables. This test can also be used for estimating the population variance. Again the test of homogeneity is applied to determine whether two or more populations are homogeneous with respect to some characteristics of interest.

- The test can be applied for dependent samples and the testing procedure is termed as "paired t test or t test for related samples." Under this approach, observations in 1st sample are related to the observations in 2nd sample.



7.7 FURTHER READING

1) Sharma, J.K. (2007). Business Statistics. New Delhi. Pearson Education Ltd.

2) Hazarika, P.L. (2016). Essential Statistics For Economics And Business Studies. New Delhi. Akansha Publishing House.

3) Lind, D.A., Marshal, W.G., Wathen, S.A. (2009) Statistical Techniques in Business and Economics. New Delhi. Tata McGraw Hill Education Private Limited.



7.8 ANSWERS TO CHECK YOUR PROGRESS

Ans to Q No 1: (i) T, (ii) F, (iii) T, (iv) T, (v) T.

Ans to Q No 2: (a) , (b) The critical value of at 5% level of significance for 3 degrees of freedom i.e., $\chi^2_{0.05,3} = 7.8147$

Ans to Q No 3: $d.f. = 3 - 1 = 2$

The critical value of at 5% level of significance for 2 degrees of freedom i.e., $\chi^2_{0.05,2} = 5.9915$

(a) The decision rule is to reject if

(b) $\chi^2 = \frac{(10-20)^2}{20} + \frac{(20-20)^2}{20} + \frac{(30-20)^2}{20} = 10.0$

(c)Reject H_0.

Ans to Q No 4: (a) Contingency table.

(b) Null hypothesis is H_0 : There is no relationship between income and whether the person played the lottery.

Alternative hypothesis is H_1 : There is a relationship between income and whether the person played the lottery.

(c) The degrees of freedom, $d.f. = (2-1)\times(3-1) = 2$

(d) Reject H_0 if $\chi^2 > 5.9915$,The critical value of χ^2 at 5% level of significance for 2 degrees of freedom i.e., $\chi^2_{0.05,2} = 5.9915$

(e)

$$\chi^2 = \frac{(46-40.71)^2}{40.71} + \frac{(28-27.14)^2}{27.14} + \frac{(21-27.14)^2}{27.14} + \frac{(14-19.29)^2}{19.29} + \frac{(12-12.86)^2}{12.86} + \frac{(19-12.86)^2}{12.86}$$

= 6.544

(f) Reject H_0.

Ans to Q No 5: Chi-square test of homogeneity is useful in a situation when we intend to verify whether several populations are homogeneous with respect to some characteristic of interest.

Ans to Q No 6: (a) Null hypothesis is $H_0 : \mu_1 = \mu_2$, Alternative hypothesis is $H_1 : \mu_1 \neq \mu_2$,

(b) Here $d.f. = 10 + 8 - 2 = 16$, Therefore tabulated value of $|t_{0.05}(16)|$ for two-tailed test = 2.120

The decision rule is to reject H_0 if $|t| > 2.120$.

(c) Pooled estimate of the population variance,

$$S^2 = \frac{(10)4^2 + (8)5^2}{10+8-2} = 22.5$$

(d) $$t = \frac{23-26}{\sqrt{22.5\left(\frac{1}{10}+\frac{1}{8}\right)}} = -1.333$$

(e) H_0 may be accepted.

Ans to Q No 7: (a) Null hypothesis is H_0: There is no change in IQ level after training programme.

(b) Reject H_0 if $t > 4.6$, the tabulated value of $|t_{0.05}(4)|$ for two-tailed test.

(c) $\bar{d} = \frac{10}{5} = 2, S = 5.477$ Therefore, $t = \frac{2}{5.477/\sqrt{5}} = 0.817$

(d) H_0 may be accepted.



7.9 MODEL QUESTIONS

Q 1: Define Chi-square distribution. Discuss the important properties of the distribution.

Q 2: Under what situations Chi-square test can be applied?

Q 3: Write down the applications of Chi-square test.

Q 4: The figures given below are (a) the theoretical frequencies of a distribution, and (b) the frequencies of the normal distribution having the same mean, standard deviation and the total frequency as in (a).

(a)	1	5	20	28	42	22	15	5	2
(b)	1	6	18	25	40	25	18	6	1

Apply χ^2-test of goodness of fit.

[Hints: Since the observed and the expected frequencies are less than 5 at the beginning and at end of the series, these frequencies are grouped

Degrees of freedom = 4, Tabulated value of $\chi^2_{0.05}(4) = 9.49$

H_0 may be accepted implying that the fit is good.]

Q 5: A Mumbai film director claims that his films are liked equally by males and females. An opinion survey of a random sample of 1000 film-goers revealed the following results:

	Liked	Disliked
Male	402	193
Female	245	160

Is the film director's claim supported by data?

[Given : $\chi^2_{0.05}$ = 3.84 5.99 7.81

d.f.= 1 2 3]

Q 6: A certain oil corporation conducts a seismic test on land to assess the possibility of extraction of oil from the land. From the past 200 records, the following test results and oil yields were obtained:

	Oil yield frequencies		
Test results	High	Medium	Low
Good	44	22	13
Fair	15	19	26
Bad	5	8	47

Can we say, at 1% level of significance that oil yields and test results are associated?

Q 7: The following table gives the classification of 100 workers according to sex and the nature of work. Test whether the nature of work is independent of the sex of the work.

	Skilled	Unskilled
Male	40	20
Female	10	30

Q 8: A certain drug was administered to 456 males out of a total of 720 in a certain locality to test its efficiency against typhoid. The incidence of typhoid is shown below. Find out the effectiveness of the drug against the disease. (The table value of χ^2 for 1 degree of freedom at 5% level of significance is 3.84)

	Infection	No infection	Total
Administering the drug	144	312	456
Without administering the drug	192	72	264
Total	336	384	720

Q 9: A sample of 20 observations gave a standard deviation of 3.72. Is this compatible with the hypothesis that the sample is drawn from a normal population with variance 4.35?

[Hints: H_0 : The sample is drawn from a normal population.
Under H_0, the test statistic $\chi^2 = \frac{ns^2}{\sigma^2}$ follows χ^2 distribution with $n-1 = 20-1 = 19\ d.f.$ Table value of $\chi^2_{0.05}(19) = 30.144.$ is rejected at 5% level.]

Q 10: The sales data of an item in six shops before and after a special promotional campaign are as under:

Shops	:	A	B	C	D	E	F
Before campaign	:	35	28	31	48	50	42
After campaign	:	58	29	30	55	56	45

Can the campaign be judged to be a success? Test at 5% level of significance?

(Significant values of $|t|$ at 1% level and 5% level for 5 d.f. respectively are 2.015 and 2.571)

*** ***** ***

UNIT 8: BASIC CONCEPTS OF RESEARCH

UNIT STRUCTURE

8.1 Learning Objectives

8.2 Introduction

8.3 Research: Meaning and Definition

8.4 Nature of Research

8.5 Types and Importance of Research

8.6 Elements of Scientific Methods

8.7 Research Ethics

8.8 Significance of Statistics in Research

8.9 Let Us Sum Up

8.10 Further Reading

8.11 Answers to Check Your Progress

8.12 Model Questions

8.1 LEARNING OBJECTIVES

After going through this unit, you will be able to-

- deduce the meaning of research and put forward a proper definition
- discuss the characteristics, objectives and motivation for research
- discuss the nature of research
- state the different types of research
- outline the importance of research
- discuss the elements of scientific methods
- deliberate on research ethics.

8.2 INTRODUCTION

This is the first unit of the Course on *Research Methodology in Economics*. This is an introductory unit and it introduces you to the field of research. Here the basic concepts of research have been discussed, viz., its meaning and definition, characteristics, objectives, nature, types, importance etc. Further, the elements/steps of the scientific approach to

research has also been examined. Finally, certain important topics on research ethics have been discussed to enable you to study the rest of the units of this course without any difficulty.

It is expected that this course will be a motivation to you to undertake research as a career. Further, please note that once you step into the field of research, always try to maintain a high degree of ethical research credibility. It is expected that this unit shall be helpful towards that direction also.

8.3 RESEARCH: MEANING AND DEFINITION

Research is a systematic and scientific inquiry to find new knowledge and question the validity of the existing knowledge. It may also be described as an activity that leads us to finding new facts, information, assisting us in verifying the available knowledge and in making us question things that are difficult to understand as per existing data.

According to (Thyer, 2001), "the word research is composed of two syllables, re and search. re is a prefix meaning again, anew or over again search is a verb meaning to examine closely and carefully, to test and try, or to probe. Together they form a noun describing a careful, systematic, patient study and investigation in some field of knowledge, undertaken to establish facts or principles."

According to Webster, 'Research is a studious inquiry or examination-critical and exhaustive investigation or experimentation having for its aim the discovery of new facts and their correct interpretation'. From this definition it can be said that research is a careful critical inquiry or examination in seeking facts and principles; diligent investigation in order to ascertain something.

O. Slesinger and M. Stephension in 'The Encyclopaedia, of social sciences define research as "the manipulation of things, concepts or symbols for the purpose of generalising to extend, correct or verify knowledge, whether that knowledge aids in construction of theory or in the practise of an art'.

According to American Sociologist Earl Robert Babbie, “Research is a systematic inquiry to describe, explain, predict, and control the observed phenomenon. Research involves inductive and deductive methods.”

Katzer, Cook & Crough have defined research as follows:

“We define research here as the systematic process of collecting and analyzing information (data) in order to increase our understanding of the phenomenon with which we are concerned or interested.”

Characteristics of Research: Based on the above definitions, the following characteristics of Research can be outlined:

- Research is directed towards the solution of a problem.
- Research is based upon observable experience or empirical evidence.
- Research demands accurate observation and description.
- Research involves gathering new data from primary sources or using existing data for a new purpose.
- Research activities are characterized by carefully designed procedures.
- Research requires expertise i.e., skill necessary to carryout investigation, search the related literature and to understand and analyse the data gathered.
- Research is objective and logical – applying every possible test to validate the data collected and conclusions reached.
- Research is all about the search of answers to unresolved issues.
- Research requires courage on the part of researcher so as to move in the uncertain and unpredictable situation.
- Research is characterized by patient and unhurried activity.
- Research is carefully recorded and reported.
- A systematic approach must be followed for accurate data. Rules and procedures are an integral part of the process that set the objective. Researchers need to practice ethics and a code of conduct while making observations or drawing conclusions.
- Research is based on logical reasoning and involves both inductive and deductive methods.
- The data or knowledge that is derived in research is in real time from

actual observations in natural settings.

- In research, there is an in-depth analysis of all data collected so that there are no anomalies associated with it.
- Research creates a path for generating new questions. Existing data helps create more opportunities for research.
- Research is analytical in nature. It makes use of all the available data so that there is no ambiguity in inference.
- Accuracy is one of the most important aspects of research. The information that a researcher obtains should be accurate and true to its nature. For example, laboratories provide controlled environment to collect data. Accuracy is measured in the instruments used, the calibrations of instruments or tools, and the final result of the experiment.

Objectives of Research: The objectives of research can be discussed under the following broad heads:

- To gain familiarity with a phenomenon or to achieve new insights to it.
- To portray accurately the characteristics of a particular individual situation or of a group.
- To determine the frequency with which something occurs or with which it is associated with something else.
- To test a hypothesis of a causal relationship between variables.

Motivations for Undertaking Research: What makes people to undertake research? Some of the answers are as follows:

- Desire to get a research degree along with its benefits.
- Desire to face the challenge in solving the unsolved problem and thereby contribute towards better social outcome.
- Desire to get intellectual joy of doing some creative and original work.
- Desire to offer service to society.
- Desire to get respectability.

8.4 NATURE OF RESEARCH

The nature of a research endeavour may be summarised as follows:

- **Systematic:** By definition, research is a systematic study. A researcher

has to follow a defined systematic procedure to conduct research.

- **Scientific:** Research follows a scientific procedure. It is worth mentioning here that the scientific procedure is being followed for ages. No one put forward it. It evolved through a process. We shall discuss the scientific approach to research in the subsequent section 1.6.
- **Objectivity:** Research is objective in nature. As an ethical practice, a researcher has to conduct his research objectively, without any personal bias.
- **Definiteness:** The outcome of research is definite. The researcher deduces certain clear outcomes at the end of his/her investigation.
- **Verifiability:** Research outcomes are verifiable in nature. What it means is that the results obtained by a researcher must be verifiable. In other words, if another researcher conducts the same study, in the same situations and with the same data set, he/she must come with similar results. This also signifies that the research activity is objective.
- **Generality:** Research also helps to arrive at certain general conclusions. Based on research findings, certain general conclusions can be derived.
- **Predictability:** Research activities are predictable in nature. In most of the cases, the researcher examines certain research hypothesis. These hypothesis signifies that the research activity are to a great extent predictable in nature. This is also valid on the ground that research is based on the previous research. And we often re-examine and re-validate the findings of the earlier researchers.
- **Modifiability:** Research findings are modifiable. What it means is that the research settings of the previous studies can be modified to arrive at more efficient, desired research outcomes.
- **Dynamicity:** Research is a dynamic process. Research is often conducted on both historical and current issues. Further, with the passage of times, newer tools and techniques are used. This significantly raises the efficiency of research.

8.5 TYPES AND IMPORTANCE OF RESEARCH

8.5.1 Types of Research

Research can be broadly classified based on its purpose, its process and its outcome.

Based on purpose, research can be classified as:

- Exploratory Research
- Descriptive Research
- Analytical Research
- Predictive Research

Based on process, research can be classified as:

- Quantitative Research.
- Qualitative Research.
- Mixed Method Research

Based on outcome, research can be classified as:

- Applied Research.
- Fundamental/ Basic/ Pure Research.
- Action Research.

We shall discuss these different types of research in some detail later in Unit 3.



CHECK YOUR PROGRESS

Q 1: Define research. Answer in about 60 words

.....

.....

.....

.....

.....

Q 2: What does objectivity of research imply? Answer in about 30 words.

.....

.....

.....

8.5.2 Importance of Research

The importance of research can be discussed under the following broad points:

- It's a tool for gaining knowledge and making learning more accessible.
- The importance of research can be visible if we see it as the only means of proving theories.
- Through research, we can gather updated information and to-the-point data.
- By differentiating between valuable and vague data, it creates credibility.
- It helps us succeed in every walk of life.
- It is a Means to Find, Gauge, and Seize Opportunities.

Let us discuss the above points in some detail.

- **Research is a tool to Acquire Knowledge Effectively:** All professions and non-professionals, including students and professors, must conduct research. It is also crucial for aspiring and experienced authors, both offline and online.

  Non-professionals who value learning will benefit from performing research since it will provide them with the world of information and skills to help them survive and enhance their lives. Finding an intriguing topic to talk and/or write about among experts and scribes, should go beyond personal experience. The purpose of conducting research can include determining what the general public wants to know or what researchers want people to realize or consider. As a result, research is critical to knowledge generation and vice versa.

  A factual proposition in a person's mind is frequently described as knowledge. It essentially refers to facts derived from objective observations and/or research findings that the human brain has processed. It can be obtained in various ways, including reading books and articles, listening to experts, watching documentaries

or investigative shows, undertaking scientific experiments, and connecting with others. Facts gathered during research can be cross-referenced with other sources to guarantee their integrity and authenticity.

- **It's used to prove theories:** Just like our brain development needs research, theoretical proofs need research, too. To understand the importance of research, we should realise that our life revolves around theories, and if we remove research from the equation, we can see that our lives are nothing but a mess. Moreover, we may derive a different point of view which we do not have to keep things to ourselves, and instead, we do it to inform humankind about what change our theory might bring. The primary purpose of all the research has kept us connected.
- **Provides the latest information:** Knowledge is one of the core items human beings have desired since the start of the human race. Being more knowledgeable and powerful has made us addicted to the latter. And this is true as well. The countries with the most knowledge base are the ones that have shown development. However, there can be other reasons for doing research. It is a well-known fact that research and development (R&D) plays a crucial role in the advancement of a country and of any organisation.
- **Research builds credibility:** Research provides a strong basis for us to construct our thoughts and beliefs on. Through proper research, we can be confident about our statement/argument. When we do our proper research on any topic, we can effectively defend our stand/point of view on it. For this, we have to do our research using the most reliable sources. On the contrary, if our research is based on opinions from non-experts, it won't be much credible. When our research is good, people are more likely to pay attention to our views/arguments.
- **Helps us succeed in every walk of life:** We need to do lot of research for undertaking an informed and effective decision.

For example, suppose you want to buy a smartphone. Before its purchase, you will definitely a great deal of research on it. The sources of such research may be a reliable website, or customer reviews, or your reliable friends who have already used such mobiles. Similarly, while opting for any course, we always conduct research on the course structure, fee structure, prospects as well as the provider and the resource persons involved (i.e., the institution). Such a research helps us protecting ourselves from wastage of our money, time and effort. The word of caution here is that in conducting such research we must use the best and most reliable resources.

Looking into a larger perspective, the public authority also has to conduct a great deal of research while designing and implementing a public policy. Only proper research can help the public authority to come out with the best and effective public policy. For example, in the implementation of the recent New Education Policy, 2020, we saw that public opinions were sought on a draft policy before finalisation of the policy.

Similarly, investment in R&D help a business organisation to gain competitive advantage. Finding ways to make things run more smoothly and differentiate a company's products from those of its competitors can help to increase a company's market worth.

- **Means to Find Gauge and Seize Opportunities:** Research also helps people build their potential and attain objectives with different opportunities. These could include gaining jobs, scholarships, education, projects, etc. For example, research is vital for those who are seeking a job or brighter pastures. Here again, through proper research, the unemployed will boost their chances of discovering potential employers through employment posts or agencies, but they need to find out if employment opportunities are authentic. Unscrupulous recruitment agencies who may be engaging in criminal recruitment and/or trafficking might be the creepy but hopeful jobseeker or migrant worker if

no inquiry is conducted. As a result, research is critical in identifying and seizing the correct opportunity in every aspect of our lives.

8.6 ELEMENTS OF SCIENTIFIC METHODS

The scientific method is a process for experimentation that is used to explore observations and answer questions. It follows therefore that do all scientists follow exactly this process? No. Some areas of science can be more easily tested than others. For example, scientists studying how stars change as they age or how dinosaurs digested their food cannot fast-forward a star's life by a million years or run medical exams on feeding dinosaurs to test their hypotheses. When direct experimentation is not possible, scientists modify the scientific method. In fact, there are probably as many versions of the scientific method as there are scientists! But even when modified, the goal remains the same: to discover cause and effect relationships by asking questions, carefully gathering and examining the evidence, and seeing if all the available information can be combined in to a logical answer.

Even though we show the scientific method as a series of steps, keep in mind that new information or thinking might cause a scientist to back up and repeat steps at any point during the process. A process like the scientific method that involves such backing up and repeating is called an iterative process.

Key Elements of Scientific Method: The scientific method describes the processes by which scientists gain knowledge about the world. It's characterized by seven key elements: questions, background research, hypotheses, experiments, observations, analyses, and conclusions.

This key elements of the scientific method can also be expressed as seven steps of research. This has been shown below.

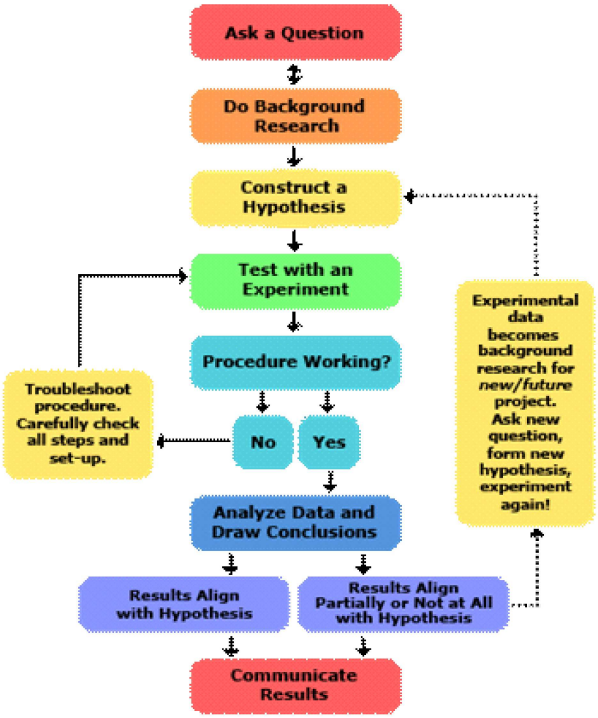
Steps in Scientific Research: The seven steps of the scientific research method/process are:

- Ask a question.
- Perform research (background research).
- Establish your hypothesis.

- Test your hypothesis by conducting an experiment.
- Make an observation (Procedure check)
- Analyse the results and draw a conclusion.
- Present the findings.

These seven steps have been shown graphically in Figure 8.1.

Figure 8.1: Key Elements/Steps of Scientific Research



In Figure 1.1, please note that the fifth phase of making an observation (or procedure check) is very crucial in the sense that we may have to move to the previous step of testing with an experiment in case any procedural lapse is noticed. Similarly, if the results arrived (in the sixth phase) are not conforming to the hypothesis, we may also have to move back again to the construction of hypothesis, i.e., to construct new hypothesis or make new experiment etc. However, this depends to a large extent on the researcher.

Now, we shall discuss these steps in some detail.

(1) Ask a Question: The scientific method starts when a researcher raises a question about something that the she/he observes: How, What, When, Who, Which, Why, or Where?

(2) Do Background Research: Rather than starting from scratch in putting together a plan for answering those questions, a researcher has to be a probing scientist using library, Internet research to find the best way to do things.

(3) Construct a Hypothesis: A hypothesis is an educated guess about how things work. It is an attempt to answer the research question with an explanation that can be tested. A good hypothesis allows the researcher make a prediction beforehand.

(4) Test the Hypothesis by Doing an Experiment: The researcher conducts experiment/tests to examine whether the prediction is accurate and thus the research hypothesis is supported or not. It is important for a researcher to conduct such experiment/test fair and free from any prejudices. The researcher can conduct a fair test by making sure that she/he change only one factor at a time while keeping all other conditions the same.

The researcher should also repeat the experiments several times to make sure that the first results weren't just an accident.

(5) Procedural check: As the testing of hypothesis conducted, the researcher also has to check that all the procedures were properly maintained. In case of any lapse in the procedure, the researcher has to repeat the experiment.

(6) Analysis of Data and Draw a Conclusion: Once the experiment is complete, the researcher has to collect the measurements and analyse them to see if they support the research hypothesis or not.

Scientists often find that their predictions were not accurate and their hypothesis was not supported, and in such cases they will communicate the results of their experiment and then go back and construct a new hypothesis and prediction based on the information they learned during their experiment. This starts much of the process of the scientific method over again. Even if they find that their hypothesis was supported, they may want to test it again in a new way.

(7) Communicate the Results: To complete the research activity, the

researcher has to communicate the results to others in a final report and/or a display board. The professional scientists do almost exactly the same thing by publishing their final report in a scientific journal or by presenting their results on a poster or during a talk at a scientific meeting. In a science fair, judges are interested in your findings regardless of whether or not they support your original hypothesis.



CHECK YOUR PROGRESS

Q 3: How does investment in R&D help an organisation? Answer in about 60 words.

.....

.....

.....

.....

.....

Q 4: Mention the different stages of scientific approach to research. Answer in about 30 words.

.....

.....

.....

8.7 RESEARCH ETHICS

Ethics generally implies those moral principles which govern an individual's behaviour and accordingly the activities performed by him or her. Webster's Dictionary defines ethical behaviour as, 'conforming to the standards of a conduct of a given profession or group'. But what is to be considered as right or wrong may vary from person to person and group to group. Ethics are important while conducting research also.

According to the Oxford Dictionary of Sociology, research ethics refers to, 'the application of moral rules and professional codes of conduct to the collection, analysis, reporting, and publication of information about research subjects, in particular active acceptance of subject's right to privacy,

confidentiality, and informed consent.'

On a broader perspective, research ethics provides guidelines for the responsible conduct of research. In addition, it educates and monitors scientists conducting research to ensure a high ethical standard. The following discussion offers a general summary of some ethical principles in research:

- **Honesty: A** researcher has to honestly report data, results, methods and procedures, and publication status. Do not fabricate, falsify, or misrepresent data.
- **Objectivity: A** researcher has to strive to avoid bias in experimental design, data analysis, data interpretation, peer review, personnel decisions, grant writing, expert testimony, and other aspects of research.
- **Integrity:** A researcher has to keep promises and agreements; act with sincerity; strive for consistency of thought and action.
- **Carefulness:** A researcher has to avoid careless errors and negligence; carefully and critically examine his/her own work and the work of the peers. A researcher also has to keep good records of research activities.
- **Openness:** A researcher should be open to share data, results, ideas, tools, resources. A researcher also should be open to receive criticisms and new ideas.
- **Respect for Intellectual Property:** A researcher has to honour patents, copyrights, and other forms of intellectual property. He/she should not use unpublished data, methods, or results without permission. A researcher also has to give credit where credit is due. Never plagiarise.
- **Confidentiality:** A researcher also has to protect confidential communications, such as papers or grants submitted for publication, personnel records, trade or military secrets, and patient records.
- **Responsible Publication:** A researcher has to publish in order to advance in research and scholarship, not to advance just in one's career. Wasteful and duplicative publication must be avoided.

- **Responsible Mentoring:** A researcher also has to educate, mentor, and advise students. The researcher must promote their welfare and allow them to make their own decisions.
- **Respect for Colleagues:** A researcher has to respect his/her own colleagues and treat them fairly.
- **Social Responsibility:** A researcher also has to strive to promote social good and prevent or mitigate social harms through research, public education, and advocacy.
- **Non-Discrimination:** A researcher must avoid discrimination against colleagues or students on the basis of sex, race, ethnicity, or other factors that are not related to their scientific competence and integrity.
- **Competence:** A researcher has to maintain and improve own professional competence and expertise through lifelong education and learning. She/he should take steps to promote competence in science as a whole.
- **Legality:** A researcher also needs to know and obey relevant laws and institutional and governmental policies.
- **Animal Care:** A researcher has to show proper respect and care for animals when using them in research. Do not conduct unnecessary or poorly designed animal experiments.
- **Human Subjects Protection:** When conducting research on human subjects, a researcher has to minimize harms and risks and maximize benefits; respect human dignity, privacy, and autonomy.

8.8. SIGNIFICANCE OF STATISTICS IN RESEARCH

Research is a systematic and scientific process which are generally carried out to acquire knowledge and information.We have already studied about the Research in earlier part of this unit. So learner are requested to acquire knowledge from that section before studying this section. Research requires investigation and interpretation of facts, phenomena and problems for the greater welbeing of society. In this regard Statistics play a vital role. In the following, we are briefly discussing the importance of Statistics in Research:

- **Collection of data:** Statistics helps in collecting the raw data from the society related to different social phenomena through the application of various tools and technique. We have already studied about the collection of data in unit 8 of this course.
- **Arrangement and presentation of data:** After collection of data, Statistics helps to arrange, process and present the data and information systematically to achieve the pre-determined social objective. We have already studied in details about the presentation of statistical data in the first course of this programme.
- **Analysis of data:** Effective arrangement of collected data helps a statistician to move to the next important step which is known as analysis of data. Here, by utilising the necessary statistical tools and technique, data is analysed to throw light on the objective of the study and establish cause and effect relationship between different variables and problems. We have already studied in details about the analysis of statistical data in the first course of this programme.
- **Interpretation of result:** Collection, presentation and analysis of data is not the sole purpose of Statistics in Research. Statistics helps the researcher to interpret the result of the study with the help of different statistical tools and technique (which we have already studied in details in unit 6 and 7 of this course), so that the result of the study is generalised and used for the betterment of the greater society.



8.9 LET US SUM UP

- Research is a systematic and scientific inquiry to find new knowledge and question the validity of the existing knowledge. It may also be described as an activity that leads us to finding new facts, information, assisting us in verifying the available knowledge and in making us question things that are difficult to understand as per existing data.
- Research can be broadly classified based on its purpose, its process and its outcome.

- The scientific method is a process for experimentation that is used to explore observations and answer questions
- Even though we show the scientific method as a series of steps, keep in mind that new information or thinking might cause a scientist to back up and repeat steps at any point during the process. A process like the scientific method that involves such backing up and repeating is called an iterative process.
- The scientific method describes the processes by which scientists gain knowledge about the world. It's characterized by seven key elements: questions, background research, hypotheses, experiments, observations, analyses, and conclusions.
- Ethics generally implies those moral principles which govern an individual's behaviour and accordingly the activities performed by him or her.
- On a broader perspective, research ethics provides guidelines for the responsible conduct of research. In addition, it educates and monitors scientists conducting research to ensure a high ethical standard.



8.10 FURTHER READING

1) Ahuja, R. (2018). *Research Methods.* Jaipur: Rawat Publications.
2) Kothari, C.R. (2011). *Research Methodology, Methods and Techniques.* New Age International (P) Ltd.
3) Kumar, R. (2011). *Research Methodology*. New Delhi: SAGE Publications India Pvt. Ltd.



8.11 ANSWERS TO CHECK YOUR PROGRESS

Ans to Q No 1: Research may be defined as an activity that leads us to finding new facts, information, assisting us in verifying the available

knowledge and in making us question things that are difficult to understand as per existing data.

Ans to Q No 2: Research is objective in nature. As an ethical practice, a researcher has to conduct his research objectively, without any personal bias.

Ans to Q No 3: Investment in R&D help a business organisation to gain competitive advantage. Finding ways to make things run more smoothly and differentiate a company's products from those of its competitors can help to increase a company's market worth.

Ans to Q No 4: The seven steps of the scientific research method/process are:

(a) Ask a question.
(b) Perform research (background research).
(c) Establish your hypothesis.
(d) Test your hypothesis by conducting an experiment.
(e) Make an observation (Procedure check)
(f) Analyse the results and draw a conclusion.
(g) Present the findings.



8.12 MODEL QUESTIONS

Answer the following questions in about 300-400 words.

Q 1: Discuss the characteristics, objectives and motivation for research.

Q 2: What do you understand by the nature of research

Q 3: Enumerate s the different types of research.

Q 4: Outline the importance of research.

Q 5: Analyse the different steps of scientific methods of research.

Q 6: Describe the important ethical issues associated the conduct of research.

*** ***** ***

UNIT 9: TYPES OF RESEARCH

UNIT STRUCTURE

9.1 Learning Objectives

9.2 Introduction

9.3 Broad Types of Research

9.4 Types of Research based on its Purpose

9.5 Types of Research based on the Process

9.6 Types of Research based on the Outcome

9.7 Conceptual and Empirical Research

9.8 Let Us Sum Up

9.9 Further Reading

9.10 Answers to Check Your Progress

9.11 Model Questions

9.1 LEARNING OBJECTIVES

After going through this unit, you will be able to:

- discuss the different types of research based on its purpose, process and outcome.
- discuss the concepts of exploratory research, descriptive research, analytical research and predictive research.
- discuss the concepts of qualitative and quantitative research and their advantages and disadvantages.
- discuss the concepts of applied research, fundamental research and action research.
- discuss the concepts of conceptual and empirical research and distinguish between them.

9.2 INTRODUCTION

In this unit, we have discussed the different types based on its purpose, process an outcome. Further, two approaches to conducting scientific research viz., conceptual research and empirical research has

also been discussed. This unit will be helpful in getting an insight into the major types of research as well as the process of research.

9.3 BROAD TYPES OF RESEARCH

Research can be broadly classified based on its purpose, its process and its outcome.

- Based on purpose, research can be classified as:
- Exploratory Research
- Descriptive Research
- Analytical Research
- Predictive Research
- Based on process, research can be classified as:
- Quantitative Research.
- Qualitative Research.
- Mixed Method Research

  Based on outcome, research can be classified as:
- Applied Research.
- Fundamental/ Basic/ Pure Research.
- Action Research.

We shall discuss these different types of research in some detail. However, we shall primarily focus on the quantitative and qualitative research types, as these are frequently used in Economics.

9.4 TYPES OF RESEARCH BASED ON ITS PURPOSE

EXPLORATORY RESEARCH: The aim of exploratory research is to look for patterns, ideas or hypotheses rather than testing or confirming a hypothesis. In exploratory research the focus is on gaining insights and familiarity with the subject area for more rigorous investigation later. It is possible that you may carry out an initial survey to establish areas of concern (exploratory research) and then research these issues in more depth, perhaps through interviews, to provide a deeper understanding (explanatory research).

As the term itself suggests, exploratory research is often conducted because a problem has not been clearly defined as yet, or its real scope is as yet unclear. It allows the researcher to familiarize him/herself with the problem or concept to be studied, and perhaps generate hypotheses to be tested. It is the initial research, before more conclusive research undertaken. Exploratory research helps determine the best research design, data collection method and selection of subjects, and sometimes it even concludes that the problem does not exist!

Exploratory research can be quite informal, relying on secondary research such as reviewing available literature, or qualitative approaches such as informal conclusions with consumers, employees, management or competitors, and more formal approaches through indepth interviews, focus groups, projective methods, case studies and pilot studies.

The results of exploratory research are not usually useful for decision making by themselves, but they can provide significant insight into a given situation. Although the results of qualitative research can give some indication as to the "why", "how" and "when" something occurs, it cannot tell us "how often" or "how many". In other words, the results can neither be generalised.

DESCRIPTIVE RESEARCH: Descriptive research describes phenomena, as they exist. It is used to identify and obtain information on the characteristics of a particular issue. It may answer such questions as:

- What is the absentee rate amongst a particular group of workers?
- What are the feelings of workers faced with redundancy?

In descriptive research, the data collected are often quantitative, and statistical techniques are usually used to summarize the information. Descriptive research goes further than exploratory research in examining a problem since it is undertaken to ascertain and describe the characteristics of issue. The methods of research utilized in descriptive research are survey methods of all kinds, including comparative and co -relational methods.

Thus, descriptive research includes surveys and fact-finding enquiries of different kinds. The major purpose of descriptive research is description of the state of affairs as it exists at present. In social science and business research we quite often use the term 'Ex post facto research'

for descriptive research studies.

The main characteristic of this method is that the researcher has no control over the variables; she/he can only report what has happened or what is happening. Most ex post facto research projects are used for descriptive studies in which the researcher seeks to measure such items as, for example frequency of shopping, preferences of people, or similar data. Ex post facto studies also include attempts by researchers to discover causes even when they cannot control the variables.

One of the major limitations of descriptive research is that it cannot help determine what causes a specific behaviour, motivation or occurrence. In other words, it cannot establish a casual research relationship between variables.

ANALYTICAL RESEARCH: Analytical research or explanatory research is a continuation of descriptive research. The researcher goes beyond merely describing the characteristics, to analyse and explain why or how something is happening. Thus, analytical research aims to understand phenomena by discovering and measuring casual relations among them. It may answer questions such as:

- How can the number of complaints made by customers be reduced?
- How can the absentee rate among employees be reduced?
- Why is the introduction of empowerment seen as a threat by departmental managers?

In analytical research, the researcher has to use facts or information already available, and analyse these to make a critical evaluation of the material. This kind of research is conducted with the assumption that a critical examination of the facts and figures will reveal more than what they appear when observed casually. So a researcher tries to discover the connection between the collected facts and figures to draw generalisations from them.

Thus, this kind of research usually tries to establish cause and effect relationship. For example, if a researcher is trying to understand a particular social context he/ she will try to discover the causes which gives rise to the events (say suicide) and the effect of those events.

However, one needs to remember that even in descriptive research some preliminary analysis can be found and similarly in the analytical researches some kind of description can be found.

PREDICTIVE RESEARCH: Predictive research goes further by forecasting the likelihood of a similar situation occurring elsewhere. It aims to generalize from the analysis by predicting certain phenomena on the basis of hypothesized general relationships. It may attempt to answer questions such as:

- Will the introduction of an employee bonus scheme lead to higher levels of productivity?
- What type of packaging will improve our products?

Predictive research provides 'how', 'why', and 'where' answers to current events as well as to similar events in the future. It is also helpful in situations where 'what if?' questions are being asked.



CHECK YOUR PROGRESS

Q 1: State the broad types of research. (Answer in about 60 words)

.....

.....

.....

.....

.....

Q 2: What is meant by exploratory research? (Answer in about 40 words)

.....

.....

.....

.....

9.5 TYPES OF RESEARCH BASED ON THE PROCESS

We have already mentioned that based on the research process, research can be categorised as qualitative research, quantitative research

and mixed method research. Here we shall discuss basically on the first two. We shall discuss the mixed method research in some detail in the next Unit 4.

QUANTITATIVE RESEARCH: The quantitative approach usually starts with a theory of a general statement proposing a general relationship between variables. With this approach it is likely that the researchers will take an objective position and their approach will be to treat phenomena as hard and real. They will favour methods such as surveys and experiments, and will attempt to test hypotheses or statements with a view to generalizing from the particular. This approach typically concentrates on measuring or counting and involves collection and analysing numerical data and applying statistical tests.

It may be defined as a systematic investigation of phenomena by gathering quantifiable data and performing statistical, mathematical, or computational techniques. Quantitative research collects information from existing and potential customers using sampling methods and sending out online surveys, online polls, questionnaires, etc., the results of which can be depicted in the form of numerical. after careful understanding of these numbers to predict the future of a product or service and make changes accordingly.

An example of quantitative research is the survey conducted to understand the amount of time a doctor takes to attend to a patient when the patient walks into the hospital. A patient satisfaction survey template can be administered to ask questions like how much time did a doctor takes to see a patient, how often does a patient walks into a hospital, and other such questions.

Quantitative outcome research is mostly conducted in the social sciences using the statistical methods statedabove to collect quantitative data from the research study. In this research method, researchers and statisticians deploy mathematical frameworks and theories that pertain to the quantity under question.

Quantitative research templates are objective, elaborate, and many times, even investigational. The results achieved from this research method

are logical, statistical, and unbiased. Data collection takes place using a structured method and conducted on larger samples that represent the entire population.

Characteristics of Quantitative Research: Some distinctive characteristics of quantitative research are:

- **Use of structured tools:** Structured tools such as surveys, polls, or questionnaires are used to gather quantitative data. Using such structure methods helps in collecting in-depth and actionable data from the survey respondents.
- **Sample size:** Quantitative research is conducted on a significant sample size that represents the target market. Appropriate sampling methods have to be used while deriving the sample to fortify the research objective
- **Close-ended questions:** Closed-ended questions are created as per the objective of the research. These questions help collect quantitative data and hence, are extensively used in quantitative research.
- **Prior studies:** Various factors related to the research topic are studied before collecting feedback from respondents.
- **Quantitative data:** Usually, quantitative data is represented by tables, charts, graphs, or any other non-numerical form. This makes it easy to understand the data that has been collected as well as prove the validity of the market research.
- **Generalization of results:** Results of this research method can be generalized to an entire population to take appropriate actions for improvement.
- **Advantages of Quantitative research:** There are many advantages of quantitative research. Some of the major advantages of why researchers use this method in market research are:
- **Collect reliable and accurate data:** As data is collected, analysed, and presented in numbers, the results obtained will be extremely reliable. Numbers do not lie. They offer an honest picture of the conducted research without discrepancies and is also extremely

accurate. In situations where a researcher predicts conflict, quantitative research is conducted.

- **Quick data collection:** A quantitative research is carried out with a group of respondents who represent a population. A survey or any other quantitative research method applied to these respondents and the involvement of statistics, conducting and analysing results is quite straightforward and less time-consuming.
- **Wider scope of data analysis:** Due to the statistics, this research method provides a wider scope of data collection.
- **Eliminate bias:** This research method offers no scope for personal comments or biasing of results. The results achieved are numerical and are thus, fair in most of the cases.

Thus, quantitative research seeks to quantify or reflect with numbers and observations about human behaviour. This kind of research attempts to describe relationships among the variables mathematically and to apply some form of numerical analysis to the relations being examined. The survey method, correlational method and experimental methods are examples of quantitative methods.

QUALITATIVE RESEARCH: Qualitative research is defined as a research method that focuses on obtaining data through open-ended and conversational communication. This method is not only about "what" people think but also "why" they think so. For example, consider a convenience store looking to improve its patronage. A systematic observation concludes that the number of men visiting this store are more. One good method to determine why women were not visiting the store is to conduct an in-depth interview of potential customers in the category.

For example, on successfully interviewing female customers, visiting the nearby stores and malls, and selecting them through random sampling, it was known that the store doesn't have enough items for women and so there were fewer women visiting the store, which was understood only by personally interacting with them and understanding why they didn't visit the store, because there were more male products than female ones.

Thus, Qualitative research is based on the disciplines of social

sciences like psychology, sociology, and anthropology. Therefore, the qualitative research methods allow for in-depth and further probing and questioning of respondents based on their responses, where the interviewer/ researcher also tries to understand their motivation and feelings. Understanding how your audience takes decisions can help derive conclusions in market research.

Qualitative research methods are designed in a manner that help reveal the behaviour and perception of a target audience with reference to a particular topic. There are different types of qualitative research methods like an in-depth interview, focus groups, ethnographic research, content analysis, case study research that are usually used.

The results of qualitative methods are more descriptive and the inferences can be drawn quite easily from the data that is obtained.

Qualitative research methods originated in the social and behavioural sciences. Today our world is more complicated and it is difficult to understand what people think and perceive. Online qualitative research methods make it easier to understand that as it is more communicative and descriptive.

Characteristics of Qualitative Research Methods: Important characteristics of qualitative research methods include:

- Qualitative research methods usually collect data at the sight, where the participants are experiencing issues or problems. These are real-time data and rarely bring the participants out of the geographic locations to collect information.
- Qualitative researchers typically gather multiple forms of data, such as interviews, observations, and documents, rather than rely on a single data source.
- This type of research method works towards solving complex issues by breaking down into meaningful inferences, that is easily readable and understood by all.
- Since it's a more communicative method, people can build their trust on the researcher and the information thus obtained is raw and unadulterated.

Advantages of Qualitative Research: Major advantaged of

qualitative research includes:

- It provides flexibility to a researcher to conduct an in depth study on a particular topic.
- The methods opted by qualitative research enables a researcher to encounter with the unknown domains of knowledge.
- Qualitative research is considered 'relatively' inexpensive.

Disadvantages of Qualitative Research: Major disadvantages of qualitative research includes:

- Qualitative research is often criticized on the ground that in such a research the researcher tends to get subjectively involved with the object of inquiry.
- Qualitative research is not replicable and therefore it is criticized on the ground that qualitative research is not often generalisable.
- Another limitation of qualitative research is that here the data collection process is more time consuming.



CHECK YOUR PROGRESS

Q 3: Define qualitative research. (Answer in about 40 words).

.....

.....

.....

.....

Q 4: Mention any two advantages of quantitative research. (Answer in about 60 words)

.....

.....

.....

.....

.....

9.6 TYPES OF RESEARCH BASED ON THE OUTCOME

We have already discussed that based on the outcome of research, it can be broadly categorised as (a) Applied, (b) Fundamental /Basic / Pure and (c) Action research. Let us briefly discuss these concepts.

APPLIED RESEARCH: Applied research is problem-oriented as the research is carried out to solve a specific problem that requires a decision, for example, the improvement of safety in the work place, or market research. Applied research is focused on answering real-world, practical questions in order to provide relatively immediate solutions. It is performed in relation to actual problems and under the conditions in which they are found in practice. Topics for applied research are often driven by current problems in education and by policymakers' concerns. Applied research is often conducted in more natural settings (i.e., more realistic or real-world settings) than basic research. Most of the educational researches are applied researches whose objectives are to develop generalizations about teaching learning processes, instructional materials, the behaviour of children, ways to modify it, and so on.

FUNDAMENTAL OR BASIC OR PURE RESEARCH: Fundamental/ basic/pure research is aimed at generating fundamental knowledge and theoretical understanding about basic human and other natural processes. The purpose of basic or fundamental research is generation of new knowledge for the sake of knowledge. It is not directed towards the solution of immediate practical problems. Basic research is usually conducted by using the most rigorous research methods (e.g., experimental) under tightly controlled laboratory conditions.

ACTION RESEARCH: Action research is focused on solving specific problems that local practitioners face in their schools and communities. It views a local classroom or other work environment as a place to conduct research. In action research, practitioners in the local settings, design and conduct the research study. It is focused more on local practice and local solutions; it is more participatory; and it is conducted by teachers,

administrators, counsellors, coaches, and other educational professionals trying to solve their very specific problems. According to Corey (1953), "Action research is the research undertaken by practitioner, so that they may improve their practices. It helps a practitioner to perceive, understand and assess the situation and it further facilitates a systematic analysis and working out plausible reasons, for the unsatisfactory condition. With action research, a practitioner can try out alternative strategies till the problem is solved satisfactorily."

9.7 CONCEPTUAL & EMPIRICAL RESEARCH

While discussing the concept of fundamental research, we mentioned that fundamental research uses very rigorous research methods like the empirical research. It should be noted here that the scientific research is often divided into two classes: conceptual research and empirical research. In this section, we shall briefly discuss the concepts of conceptual and empirical research.

CONCEPTUAL RESEARCH: Conceptual research focuses on the concept or theory that explains or describes the phenomenon being studied. For example:

- What causes disease?
- How can we describe the motions of the planets?
- What are the building blocks of matter?

The conceptual researcher sits at his/her desk with pen in hand and tries to solve these problems by thinking about them. She/he does no experiments but may make use of observations by others, since this is the mass of data that she/he is trying to make sense of.

Conceptual research typically involves observing and analysing information already present on a given topic. Philosophical research is a generally good example for conceptual research. It can be used to solve real-world problems. Conceptual frameworks, which are analytical tools researchers use in their studies, are based on conceptual research. Furthermore, these frameworks help to make conceptual distinctions and organize ideas the researchers need for research purposes.

In simple words, a conceptual framework is the researcher's synthesis of the literature (previous research studies) on how to explain a particular phenomenon. It explains the actions required in the course of the study based on the researcher's observations on the subject of research as well as the knowledge gathered from previous studies.

EMPIRICAL RESEARCH: Empirical research is basically a research technique that uses empirical evidence. Empirical evidence refers to evidence verifiable by observation or experience rather than theory or pure logic. Thus, empirical research is research studies with conclusions based on empirical evidence. Moreover, empirical research studies are observable and measurable.

Empirical evidence can be gathered through qualitative research studies or quantitative research studies. Qualitative research methods gather non-numerical or non-statistical data. Thus, these type of studies helps to understand the underlying reasons, opinions, and motivations behind something as well as to uncover trends in thought and opinions. Quantitative research studies, on the other hand, gather statistical data. These have the ability to quantify behaviours, opinions, or other defined variables. Moreover, a researcher can even use a combination of quantitative and qualitative methods to find answers to his research questions.

Distinctions between Conceptual and Empirical Research: The following broad distinctions between conceptual and empirical research can be noted.

- **Based on definition:** Conceptual research is a type of research that is generally related to abstract ideas or concepts whereas empirical research is any research study where conclusions of the study are drawn from evidence verifiable by observation or experience rather than theory or pure logic.
- **Based on nature:** Conceptual research involves abstract idea and concepts; however, it doesn't involve any practical experiments. Empirical research, on the other hand, involves phenomena that are observable and measurable.
- **Based on type of studies:** Philosophical research studies are examples of conceptual research studies, whereas empirical

research includes both quantitative and qualitative studies.

To sum up, we can say that the main difference between conceptual and empirical research is that conceptual research involves abstract ideas and concepts whereas empirical research involves research based on observation, experiments and verifiable evidence.



CHECK YOUR PROGRESS

Q 5: Define action research. (Answer in about 60 words)

.....

.....

.....

.....

.....

Q 6: Mention any two situations when applied research can be undertaken. (Answer in about 30 words)

.....

.....

.....



9.8 LET US SUM UP

- Research can be broadly classified based on its purpose, its process and its outcome.
- Based on purpose, research can be classified as: exploratory research, descriptive research, analytical research and predictive research.
- Based on process, research can be classified as quantitative, qualitative and mixed method research.
- Based on outcome, research can be classified as applied, fundamental or basic or pure and action research.
- In exploratory research the focus is on gaining insights and familiarity

with the subject area for more rigorous investigation later.

- Descriptive research describes phenomena, as they exist. It is used to identify and obtain information on the characteristics of a particular issue.
- Analytical research or explanatory research is a continuation of descriptive research. The researcher goes beyond merely describing the characteristics, to analyse and explain why or how something is happening. Thus, analytical research aims to understand phenomena by discovering and measuring casual relations among them.
- Predictive research goes further by forecasting the likelihood of a similar situation occurring elsewhere. It aims to generalize from the analysis by predicting certain phenomena on the basis of hypothesized, general relationships,
- The quantitative approach usually starts with a theory of a general statement proposing a general relationship between variables. With this approach, it is likely that the researchers will take an objective position and their approach will be to treat phenomena as hard and real.
- Qualitative research is defined as a research method that focuses on obtaining data through open-ended and conversational communication. This method is not only about "what" people think but also "why" they think so.
- Applied research is problem-oriented as the research is carried out to solve a specific problem that requires a decision, for example, the improvement of safety in the work place, or market research. Applied research is focused on answering real-world, practical questions in order to provide relatively immediate solutions.
- Fundamental/basic/pure research is aimed at generating fundamental knowledge and theoretical understanding about basic human and other natural processes. The purpose of basic or fundamental research is generation of new knowledge for the sake of knowledge.
- Action research is focused on solving specific problems that local practitioners face in their schools and communities. It views a local

classroom or other work environment as a place to conduct research. Conceptual research focuses on the concept or theory that explains or describes the phenomenon being studied.

- Empirical research is basically a research technique that uses empirical evidence. Empirical evidence refers to evidence verifiable by observation or experience rather than theory or pure logic.



9.9 FURTHER READING

1) Ahuja, R. (2018). Research Methods. Jaipur: Rawat Publications.

2) Kothari, C.R. (2011). Research Methodology, Methods and Techniques. New Age International (P) Ltd.

3) Kumar, R. (2011). Research Methodology. New Delhi: SAGE Publications India Pvt. Ltd.



9.10 ANSWERS TO CHECK YOUR PROGRESS

Ans to Q No 1: Research can be broadly classified based on its purpose, its process and its outcome. Based on purpose, research can be classified as: exploratory research, descriptive research, analytical research and predictive research. Based on process, research can be classified as quantitative, qualitative and mixed method research. And, based on outcome, research can be classified as applied, fundamental or basic or pure and action research.

Ans to Q No 2: Exploratory research may be defined as the research approach that looks for patterns, ideas or hypotheses rather than testing or confirming a hypothesis. In exploratory research the focus is on gaining insights and familiarity with the subject area for more rigorous investigation later.

Ans to Q No 3: Qualitative research is defined as a research method that focuses on obtaining data through open-ended and conversational

communication. This method is not only about "what" people think but also "why" they think so.

Ans to Q No 4: Two important advantages of quantitative research are:

- As data is collected, analysed, and presented in numbers, the results obtained will be extremely reliable. Data offer an honest picture of the conducted research without discrepancies and is also extremely accurate. In situations where a researcher predicts conflict, quantitative research is conducted.
- This research method offers no scope for personal comments or biasing of results. The results achieved are numerical and are thus, fair in most cases.

Ans to Q No 5: Action research is the research undertaken by practitioner, so that they may improve their practices. It helps a practitioner to perceive, understand and assess the situation and it further facilitates a systematic analysis and working out plausible reasons, for the unsatisfactory condition. With action research, a practitioner can try out alternative strategies till the problem is solved satisfactorily,

Ans to Q No 6: Most of the educational researches are applied researches whose objectives are to develop generalizations about teaching learning processes, instructional materials, the behaviour of children, ways to modify it, and so on.



9.11 MODEL QUESTIONS

Answer the following questions.

Q 1: Discuss the different types of research.

Q 2: Discuss the different types of research based on its purpose.

Q 3: Discuss the different types of research based on the process.

Q 4: Discuss the different types of research based on the outcome.

Q 5: Discuss the concepts of conceptual and empirical research.

Q 6: Discuss the relative merits of qualitative and quantitative research.

Q 7: Briefly discuss the concept of exploratory research.

*** ***** ***

UNIT 10: REVIEW OF LITERATURE

UNIT STRUCTURE

10.1 Learning Objectives

10.2 Introduction

10.3 Review of Literature: Meaning and Concept

10.3.1 The 5 C's of a Literature Review

10.4 Need for Reviewing Literature

10.5 Sources of Literature

10.6 Let Us Sum Up

10.7 Further Reading

10.8 Answer to Check Your Progress

10.9 Model Questions

10.1 LEARNING OBJECTIVES

After going through this unit, you will be able to:

- discuss the concept and meaning of review of literature
- discuss the 5 Cs of literature review
- outline the need for reviewing literature
- describe the various sources of literature.

10.2 INTRODUCTION

When we share knowledge it multiplies. In order to share the acquired, obtained or generated knowledge it is necessary that it must be transmitted, preserved and advanced. But advancement in knowledge takes place when there is suitable transmission and preservation. Practically knowledge can be preserved and transmitted through books, libraries and educational entities. For advancement of knowledge, it is necessary to review the existing knowledge and improve it on logical and justifiable premises. Logic and justification always stem from investigation. But before entering into any investigation an investigator must ensure that the enquiry is not from a vacuum; there should be considerable work to pave the way. Appraising existing scholarly work is called review of literature. Sometimes,

an examination into earlier scholarly work itself provides solution to the problem or answer to a question. In any discipline for the purpose of advancement and up-gradation, a researcher must be thoroughly familiar with both previous theory and hypothesis. In each of the behavioural sciences research, review of literature starts at the very initial stage of the whole research plan and continues till its completion.

In this unit, we shall discuss the concept and meaning of review of literature, its need and various sources.

10.3 REVIEW OF LITERATURE: MEANING AND CONCEPT

Dissection of the term "Review of literature" is self explanatory, it consists of two words i.e., review + literature. The dictionary meaning of review is "a look back at something in order to check, remember, or be clear about something" or "the examining or considering over again of something in order to decide if transformation is needed". Similarly, the dictionary meaning of literature is "pieces of writing that are valued as works of art, especially novels, plays and poems" or "pieces of writing or printed information on a particular subject".

The expression "review of literature" is basically an examination of printed information on a particular subject or a matter. Printed information can be sourced from books, research articles, scholarly work, policy papers, monographs, magazines, newspapers, scriptures, and Government reports, etc.

W.R. Borg has defined review of literature as:

[The literature in any field forms the foundation upon which all future work will be built. If we fail to build the foundation of knowledge provided by the review of literature our work is likely to be shallow and naive and will often duplicate work that has already been done better by some one else.]

Good, Barr and Scates have defined it as:

[The competent physician must keep abreast of the latest discoveries in the field of medicine. Obviously the careful student of education, the research worker and investigator should become familiar with location and use of

sources of educational information.]

Appraising literature is crucial for thesis and any research work. It takes place in two phases. The first phase is called identification phase and the second phase is called foundation stage. In the first phase of the literature-review a researcher identifies all the published material which is related to the research problem and reads those parts with which the researcher is not familiar. It is then that the foundation of the study or formulation of research work takes place. In the second phase, the related material is penned down by the researcher and organised in the relevant section in order to justify or support the thesis/ research work. This particular section in which the literature review is presented engages the attention of both reader as well as researcher. The researcher justifies the background of the study based upon it whereas the readers come to know about earlier researches in the area which develops the understanding for the study. In most of the behavioural sciences, review of theoretical and empirical studies are executed by the researchers. In this manner a complete theoretical foundation with proper justification is laid by the researcher. Hence, it can be said that review of existing literature is an inseparable part of an examination which is to be made or has been performed by the investigator (researcher).

Thus, we can say that literature review is a survey of scholarly sources (such as books, journal articles, and theses) related to a specific topic or research question. It is often written as part of a thesis, dissertation, or research paper, in order to situate the work of the researcher in relation to existing knowledge.

Different parts of literature reviews: Just like an academic paper, literature reviews also contain at least three basic elements: an introduction or background information section; the body of the review containing the discussion of sources; and, finally, a conclusion and/or recommendations section to conclude the discussion.

10.3.1 The 5 C's of a Literature Review

Many researchers have discussed 5 Cs of literature review.

These 5 Cs help a researcher to undertake the review of literature systematically. These 5 Cs are: Concise, Clear, Critical, Convincing and Contributing. In the following paragraphs, we shall discuss each of these in some detail.

- **Concise:** The amount of literature available on any given topic can be vast. In today's IT age, the problem becomes even more severe. This is because today one finds thousands of articles and papers on a given topic whenerver an internet search isa made. However, the job of a researcher here is to consciously undertake the literature review based primarily on the research questions. It is at this early stage that the researcher will need to decipher where the gaps of knowledge are within the current research. It has to be narrowed down and presented in a crisp and concise manner.
- **Clear:** The literature pertaining to a research topic will also vary. Some material will be conceptual and other material will be empirical. Referring back to the research questions and asking how the literature pertains to the topic is essential. This can help the researcher to identify common themes, threads and patterns which will help bring about clarity about the topic under study.
- **Critical:** A literature review is more than a summary of materials. Each piece of literature should be critically analyzed for assumptions, beliefs, and values that essentially influence the authors' thought. As a researcher it is at this point that one needs to be critical of the works of others and evaluate the underlying assumptions that may have guided the available works.
- **Convincing:** Synthesis of the literature demands that the researcher gathers all the analyses and develops convincing arguments. The basic goal of the researcher here is to arrive at a valid interpretation of literature resulting from a critical analysis combined with supporting data.
- **Contributing:** Inclusion of the most relevant information is essential to the literature review. The goal of a literature review

is to support research where it is required. The review is intended to extend current knowledge, theory and be influential. The review is not a summary of previous work but a new branch created and supported by the researcher. Thus, it should always try to contribute newer insights into the current realm of research.

10.4 NEED FOR REVIEWING LITERATURE

The review of literature is essential because of the following reasons:

- **Planning and execution of the study:** The manner in which the research work will be executed can be decided by looking into the available past work. Up to date knowledge of the research findings makes it clear aboutb what has already been done and how it has been executed. Based upon earlier work, a plan can easily be chalked out for the problem in hand in a constructive manner.
- **Minimization of replication and repetition:** Awareness helps us to avoid repetition of the same work. Good research work is expected to contain an element of novelty. Hence, review of related literature nullifies replication of the same work.
- **Direction to the study:** Precedent research work by varicious scholars provides insight into the nature of the existing problem and outlines a direction for the study. Quantitative as well as qualitative analysis of the previous work of other scholars also suggests the methodology to attain the objectives of the study.
- **Formulation of Hypothesis:** Hypotheses are assumptions formulated by the researcher in context of a particular phenomenon. But hypotheses cannot be constructed merely on a whim or a gut feeling. Literature survey is an aid to develop a hypothesis. Prominent work in the area provides rationale for the supposition. Approval and disapproval of suppositions made a researcher to discuss the results of a study at length duly supported by the test and data .
- **Selection of variable and functional form:** In behavioural sciences research, most of the variables are supported by earlier research works. It makes selection of variables less controversial and more

authenticated. Functional relationship among the variables can easily be figured out with the help of through reading. Theoretical foundations always pave a way for the empirical testing. For example, the theory of demand holds that there is an inverse relationship between the price and quantity demanded. This functional relationship can be verified in context of various consumer and industrial goods.

- **Identification of research Gap:** Research gap is an issue, query, question or dimension which has not been raised, solved, answered or explored by any of the existing studies or investigations. How can one come to know that a particular dimension of a problem has not been explored till the time? It is only a literature survey that can make it clear. Before a researcher enters into formulation of research objective or hypothesis development of a literature survey is a valuable asset.



CHECK YOUR PROGRESS

Q 1: What is meant by literature review? Answer in about 50 words.

.....

.....

.....

.....

.....

Q 2: Why is review of literature important? Answer in about 30 words.

.....

.....

.....

10.5 SOURCES OF LITERATURE

The major sources of literature for review includes the following:

- **Reference Books and Text books:** A large number of books are

published in English. The list of such books can be sourced from the publisher in the form of a catalogue or it can be fetched from the libraries. Besides, the website of publishers contains lists of books as well as table of contents of the books published by a particular publication house.

- **Periodicals:** These are printed material which are published regularly after a particular period of time. The time lag is defined, even and distributed over a period of a year. In other words it can be said that periodicals are expected to continue for infinite time but at a regular interval. For example, *Economic and Political Weekly* is a research periodical published in India. Periodicals include Yearbooks, Documents, Almanacs, The Cumulative Book Index, International Abstracts, Journals, Newspapers, Magazines, International Index to Periodicals etc.
- **Abstracts:** It presents the summary of the content. It presents a brief summary of the research work. It also presents the precise idea of the research work. A wide variety of indexes and general references can be found to cover almost any area in which the modern researcher might be interested. In addition to it, the electronic data base that contains the research papers provides free access to the abstracts of the scholarly work.
- **Encyclopaedias:** Encyclopaedias provide concise data on various subjects composed by trained professionals. For specialized field of knowledge there are specialized Encyclopaedias. These contain illustrations, examples and bibliographies.
- **Almanacs, Handbooks, Yearbooks and Guides:** Handbooks and Yearbooks contain information which is statistical in nature. For example SEBI handbook which contains all the quantitative data related to the stock market. Almanacs and guides are general category of reference that includes publications that contain detailed up-to-date information on various subjects and are well thought out according to a particular theme.

'World Almanac' Book of Facts, New York: It is source of

miscellaneous information of various subjects. 'Handbook of Research on Teaching' Chicago: It provides comprehensive research on teaching within depth and extensive bibliographies.

- **Specialized Dictionaries:** These dictionaries contain the meaning of subject specific terminology, expression and jargon. For example Oxford dictionary of English Language, Oxford Dictionary of Finance, Macmillan Dictionary of Economics, Routledge Dictionary of Economics etc.
- **ERIC (Educational Research Information Centre):** The current knowledge explosion makes selective data retrieval the key to the research enterprise as well as to effective educational practice. Major developments in this regard as they relate to the educational literature are ERIC (Educational Research Information Centre) and SRIS (School Research Information Service). ERIC is an attempt to facilitate information exchange and to increase the value of research to the educational community by simply making its results readily available in usable form. In our country NCERT has established a separate ERIC cell to facilitate educational research community
- **Microfiche:** "A microfiche is a sheet of film containing micro-images of printed material. A copy of film 4" × 6" card contains the material of one hundred printed pages of 9" × 11" size. There are many document reproduction services that supply microfiche to libraries upon special order. Super-and Ultra Microfiche is a recent development in the field of micro-printing. It has transformed the process of storage of published material in libraries of the future. A super microfiche has been developed that contains up to two pages of printed material on a single 4" × 6" transparent card, the equivalent of two or more books. An even more spectacular development is the ultra- microfiche that contains up to 3200 micro-dots on a single card of 4" × 6". When projected each dot contains the equivalent of several pages. Thus, seven to ten volumes can be included on a single 4" × 6" transparent card. Reader printers make hard copy prints of any page in few seconds".

- **Dissertations and Theses:** Theses and dissertations which embody the bulk of educational research are usually housed by institutions and universities that award authors their advanced degrees. Sometimes these studies are published in whole or in part in educational journals. The related dissertations and theses are the main sources of review of literature. The entry 'dissertations and theses' issue of the bibliographic index is the most comprehensive listing of sources of these research in progress.
- **Other Sources:** Besides all these sources following are also important sources of available literature:
  - Newspaper
  - Conference papers
  - Electronic Data-base for example emerald, proquest, J-stor etc.
  - Government reports



CHECK YOUR PROGRESS

Q 3: Mention the major sources for literature review. Answer in about 30 words.

.....

.....

.....

Q 4: What is ERIC? Answer in about 30 words.

.....

.....

.....



10.6 LET US SUM UP

- In order to share the acquired, obtained or generated knowledge it is necessary that it must be transmitted, preserved and advanced.
- Advancement in knowledge takes place when there is suitable transmission and preservation. Knowledge can be preserved and

transmitted through books, libraries and educational entities.

- For advancement of knowledge, it is necessary to review the existing knowledge and improve it on logical and justifiable premises.
- Before entering into any investigation an investigator must ensure that the enquiry is not in a vacuum; there should be considerable work to pave the way.
- Appraising existing scholarly work is called review of literature.
- On several occassions an examination into earlier scholarly work itself provides solution to the problem or answer to a question.
- In any discipline for the purpose of advancement and up-gradation, a researcher must be thoroughly familiar with both previous theory and hypothesis.
- Many researchers have discussed 5 Cs of literature review. These 5 Cs help researchers to undertake the review of literature very systematically. These 5 Cs are: Concise, Clear, Critical, Convincing and Contributing.
- Appraising literature is crucial for thesis and any research work. It takes place in two phases.
- The first phase is called identification phase and the second phase is called foundation stage. In the first phase of the literature-review a researcher identifies all published material which is related to the problem and reas related materials with which the researcher is not familiar and the foundation of the study or formulation of research work begins.
- In the second phase the related material is organised by the researcher and placed in the relevant section in order to justify or support the thesis/ research work.
- This particular section in which the literature is presented fetches the attention of both reader as well as researcher. The researcher justifies the background of the study based upon it whereas the readers come to know about the abstract and earlier research which develop the understanding for the study.
- Review of literature is essential for planning and execution of the

study, minimization of replication and repetition, giving direction to the study, formulation of hypothesis, selection of variables and functional form etc.

- The major sources of literature for review include books and text books, periodicals, abstracts, encyclopaedias, almanacs, handbooks, yearbooks and guides, specialized dictionaries, ERIC, microfiche etc.
- Apart from those, other important sources include: newspapers, conference paper, electronic data-base for example emerald, pro-quest, J-stor, government reports etc.



10.7 FURTHER READING

1) Cooper D. and Schindler P. Business Research Methods, Tata McGraw Hill.

2) Goode and Heart, Research Method

3) Jerry W. Willis, Foundations of Qualitative research: Interpretative and Critical Approaches. Sage.

4) Kothari R.C. Research Methodology, Methods and Techniques, New Age International Publishers, 2nd revised edition, Reprint 2008.



10.8 ANSWERS TO CHECK YOUR PROGRESS

Ans to Q No 1: Literature review is a survey of scholarly sources (such as books, journal articles, and theses) related to a specific topic or research question. It is often written as part of a thesis, dissertation, or research paper, in order to situate the work of the researcher in relation to existing knowledge.

Ans to Q No 2: Review of literature is needed for planning and execution of the study, minimization of replication and repetition, direction to the study, formulation of hypothesis. selection of variables and functional

form etc.

Ans to Q No 3: Major sources of literature for review include books and text books, periodicals, abstracts, encyclopaedias, almanacs, handbooks, yearbooks and guides, specialized dictionaries, ERIC, microfiche etc.

Ans to Q No 4: ERIC is an attempt to facilitate information exchange and to increase the value of research to the educational community by simply making its results readily available in usable form.



10.9 MODEL QUESTIONS

Answer the following questions in about 300 words.

Q 1: Discuss the concept and meaning of review of literature.

Q 2: Discuss the need for reviewing the existing literature.

Q 3: Discuss the various sources of literature review.

Q 4: Discuss the 5 C's of literature review.

*** ***** ***

UNIT 11: RESEARCH PROBLEM

UNIT STRUCTURE

11.1 Learning Objectives

11.2 Introduction

11.3 Identification of the Research Problem

11.4 Selection of Research Problem: Contributing Factors

11.5 Technique Involved in Defining a Problem

11.6 Let Us Sum Up

11.7 Further Reading

11.8 Answer to Check Your Progress

11.9 Model Questions

11.1 LEARNING OBJECTIVES

After going through this unit, you will be able to:

- identify a research problem
- discuss the factors involved in the selection of a research problem
- describe the technique involved in the selection of a research problem

11.2 INTRODUCTION

This unit deals with the discussion on research problem. In the process of research, proper definition and identification of the research problem is of utmost importance. The entire outcome of research depends on the identification of the research problem. How we identify the research problem and how best we describe it, is very crucial for all other research activities, viz., formulation of research hypothesis, application of research tools etc. This unit will be helpful in proper identification of the research problem as well as in selection of research problem and utilization of proper technique in that process.

11.3 IDENTIFICATION OF THE RESEARCH PROBLEM

In research process, the first and foremost step happens to be that of identifying and properly defining a research problem. Identification of research problem is a brainstorming process and it involves thinking in a logical manner. Normally, thinking may be classified into four categories: convergent, divergent, reflective and scientific thinking. Convergent thinking basically focuses on coming up with the single, well-established answer to a problem. It is oriented toward deriving the single best, or most often the correct answer to a question. Convergent thinking is used as a tool in creative problem solving. On the other hand, divergent thinking is described as the thought process or method used to generate creative ideas by exploring many possible solutions. It typically occurs in a spontaneous, free-flowing, "non-linear" manner, such that many ideas are generated in an emergent cognitive fashion. Similarly, in reflective thinking individual conceived for a new solution for an old problem, but scientific thinking is in terms of carefully organized reflection. Identification of a research problem is exploratory in nature and it is sought to understand the pattern of relationship between/ among different variables. A researcher must find the problem and formulate it so that it becomes susceptible to research. To identify a problem correctly, a researcher must know: what is a problem? The research problem can be identified either in qualitative or quantitative terms. In a simple terms identification of pattern of occurrence or what, why and how, of a happening is called identification of research problem whose solution or answer is sought by the researcher. Following are the crucial elements of a research question or a research problem:

There should be an entity or entities which needs an investigation or enquiry. sThere should be some research goals to be achieved. If one does not have a research goals, one will not know what research problem to pursue. One should have access to resources and methods to attain the research goals.. If there will not be alternative methods to attain the defined research goals, the researcher will not have any problem to be identified.

There should be some sort of rationality with respect to selection of

methods and means. It will make a researcher to justify the relevance of a particular method and means.

There should be some background to which the trouble/ problem/ ambiguity relates. It could be an event or a socio-economic and psychological phenomenon.

For example, if an economist has to explore the reasons of fall in the gross domestic product, it is a problem related to individual inquisitiveness. What the economist will attain by figuring out the reasons of low gross domestic product? - It is related to the objective of the economist. Through which particular methods the economist will be evidencing his findings? It is related to selection of available method and means.

Thus, a research problem is one which requires a researcher to find out the best solution for the given problem, i.e., to find out by which course of action the objective can be attained optimally in the context of a given environment.

11.4 SELECTION OF RESEARCH PROBLEM: CONTRIBUTING FACTORS

Several factors can complicate identification of a research problem, such as, large number of -interconnected events or socio-economic phenomena, personal bias, lack of resources, delayed response, dynamic environment, technology transformation, and response of the groups or individuals. All or some of these factors may be crucial to identify a research problem. Nevertheless, following factors might be involved in the selection of problem.

- **Individual preference and interest:** Selection of a research problem can be made by a corporate or a non-corporate entity. A corporate entity looks for solution of the problem that can bring in revenue whereas, a non-corporate entity always picks up the issues related to the society, people and environment. In addition to it a person tries to address the research problem according to inclination and personal interest. Chief motivation in selection of a research problem is the

personal inclination of the researcher. For example, the person who has interest in agriculture economics will abstain from the issues of business economics.

- **Time-lines and urgency:** Time- line is the duration in which solution to a particular problem/ situation or answer to a specific question is expected. The nature of the problem and urgency are the deciding factors in selection of a problem. For example, the time period required to diagnose the reason of hair fall is less in contrast to the time period required to develop a new medicine for it. In the similar manner, vaccine for COVID developed in a very short duration in comparison to the time period required to develop a vaccine in the normal situation due to the urgency factor. In some cases, there may be certain issues that are exceptionally captivating, promising, and have importance, yet the researcher cannot examine them in the given time.
- **Resources accessibility:** In order to address a particular research issue availability of resources is significant. In behavioral sciences books, magazine, library facilities etc. are significant in selection of a research topic. For example, it will be difficult to take up the issue of national interest as a research project for a person who is there in a college in some remote area due to limited availability of resources like library, database, software for data analysis etc. Whereas if a person is in some university, the resources are easily available, and consequently selection of a research problem or research topic vary in quality and nature.

  In core sciences or engineering in addition to library resources the availability of laboratory, equipment and materials are crucial for selection of a research topic.
- **Importance and significance:** In behavioral sciences the issues which are significant in practical life is additionally answerable by the choice of a research topic. The general surrounding also contributes in selection of a research problem. Everybody appreciates pursuing a research objective that has a great deal of pragmatic ramifications. The importance of a particular issue to the society and availability of

resources leads to the selection of the research problem. In addition, selection of the topic should be such that it promises a solution.

- **Researcher Knowledge:** Understanding of the problem becomes better when it is backed by the knowledge component. Skills and competencies of a researcher can aid the knowledge of the researcher.

  Researcher's subject specific knowledge and depth of information should can assist the researcher in effectively choosing a research problem that requires investigation.



CHECK YOUR PROGRESS

Q 1: Define a research problem. Answer in about 50 words.

.....

.....

.....

.....

11.5 TECHNIQUE INVOLVED IN DEFINING A PROBLEM

Let us start with the question: What does one mean when he/she wants to define a research problem? The answer may be that one wants to state the problem along with the bounds within which it is to be studied. In other words, defining a problem involves the task of laying down boundaries within which a researcher shall study the problem keeping in view a pre-determined objective.

Defining a research problem is undoubtedly a herculean task. However, it must be tackled intelligently to avoid the ambiguity encountered in a research operation. The usual approach is that the researcher should himself/herself pose a question (or in case someone else wants the researcher to carry on research, the concerned individual, organisation or an authority should provide a problem/question to the researcher) and set-up techniques and procedures for throwing light on the question concerned for formulating or defining the research problem. But such a casual approach

generally does not produce definitive results because the question framed in such a fashion may be very broad general terms and as such may not be in a form suitable for testing.

Defining a research problem properly and clearly is a crucial part of a research study and must in no case be accomplished hurriedly. However, in practice this is frequently overlooked which may land a researcher into problems later on. Hence, the research problem should be defined in a systematic manner, giving due weightage to all related points. The technique for the purpose involves undertaking the following steps: (i) statement of the problem in a general way; (ii) understanding the nature of the problem; (iii) surveying the available literature (iv) developing the ideas through discussions; and (v) rephrasing the research problem into a working proposition.

A brief description of all these points will be helpful.

- **Statement of the problem in a general way:** First of all, the problem should be stated in a general way, keeping in view either some practical concern or some scientific or intellectual interest. For this purpose, the researcher must immerse himself/herself thoroughly in the subject matter concerning which he/she wishes to pose a problem. In social research, it is advisable to do some field observation and as such the researcher may undertake some sort of preliminary survey or what is often called pilot survey. Then the researcher can himself/herself state the problem or he/she can seek the guidance of the guide or the subject expert in accomplishing this task. Often, the guide puts forth the research problem in general terms, and it is then up to the researcher to narrow it down and phrase the problem in operational terms. In case there is some directive from an organisational authority, the problem then can be stated accordingly. The problem stated in a general way may contain various ambiguities which must be resolved by systematically thinking and rethinking over the problem. At the same time the feasibility of a particular solution has to be considered and the same should be kept in view while stating the problem.
- **Understanding the nature of the problem:** The next step in

defining the problem is to understand its origin and nature clearly. The best way of understanding the problem is to discuss it with those who first raised it in order to find out how the problem originally came about and with what objectives in view. If the researcher has stated the problem himself/herself, he/she should consider once again all those points that induced him/her to make a general statement concerning the problem. For a better understanding of the nature of the problem involved, he/she can enter into discussion with those who have a good knowledge of the problem concerned or similar other problems. The researcher should also keep in view the environment within which the problem is to be studied and understood.

- **Review of literature:** All available literature concerning the problem at hand must necessarily be reviewed and examined before a definition of the research problem is given. This means that the researcher must be well-conversant with relevant theories in the field, reports and records, as also all other relevant literature. He/she must devote sufficient time in reviewing of research already undertaken on related problems. This is done to find out what data and other materials, if any, are available for operational purposes. Proper knowledge about the data available often serves to narrow the problem itself as well as the technique that might be used. This would also help a researcher to know if there are certain gaps in the theories, or whether the existing theories applicable to the problem under study are inconsistent with each other, or whether the findings of the different studies do not follow a pattern consistent with the theoretical expectations and so on. All this will enable a researcher to take new strides in the field for furtherance of knowledge i.e., he can move up starting from the existing premise. Studies on related problems are useful for indicating the type of difficulties that may be encountered in the present study as also the possible analytical shortcomings. At times such studies may also suggest useful and even new lines of approach to the present problem.

- **Developing the ideas through discussions:** Discussion concerning a problem often produces useful information. Various new ideas can be developed through such an exercise. Hence, a researcher must discuss his/her research problem with his/her colleagues and others who have enough experience in the same area or is working on similar problems. This is quite often known as an experience survey. People with rich experience are in a position to enlighten the researcher on different aspects of his/her proposed study and their advice and comments are usually invaluable to the researcher. They help him/her sharpen his focus of attention on specific aspects within the field. Discussions with such persons should not only be confined to the formulation of the specific problem at hand, but should also be concerned with the general approach to the given problem, techniques that might be used, possible solutions, etc.
- **Rephrasing the research problem:** Finally, the researcher must sit to rephrase the research problem into a working proposition. Once the nature of the problem has been clearly understood, the environment (within which the problem has got to be studied) has been defined, discussions over the problem have taken place and the available literature has been surveyed and examined, rephrasing the problem into analytical or operational terms is not a difficult task. Through rephrasing, the researcher puts the research problem in as specific terms as possible so that it may become operationally viable and may help in the development of working hypotheses.

  In addition to what has been stated above, the following points must also be observed while defining a research problem:

  - Technical terms and words or phrases, with special meanings used in the statement of the problem, should be clearly defined.
  - Basic assumptions or postulates (if any) relating to the research problem should be clearly stated.
  - A straight forward statement of the value of the investigation (i.e., the criteria for the selection of the problem) should be

provided.

- The suitability of the time-period and the sources of data available must also be considered by the researcher in defining the problem.
- The scope of the investigation or the limits within which the problem is to be studied must be mentioned explicitly in defining a research problem.



CHECK YOUR PROGRESS

Q 2: Mention the four major steps in defining a research problem. Answer in about 50 words.

.....

.....

.....

.....

.....

Q 3: How does review of literature help in the research process. Answer in about 60 words.

.....

.....

.....

.....

.....

.....



11.6 LET US SUM UP

- The task of defining a research problem, very often, follows a sequential pattern-the problem is stated in a general way, the ambiguities are resolved, thinking and rethinking process results in a more specific formulation of the problem so that it may be a realistic one in terms of the available data and resources and is also

analytically meaningful.

- All this results in a well-defined research problem that is not only meaningful from an operational point of view, but is equally capable of paving the way for the development of working hypotheses and for means of solving the problem itself.
- Defining a research problem properly and clearly is a crucial part of a research study and must in no case be accomplished hurriedly. However, in practice this is frequently overlooked causing problems later in research.
- The research problem should be defined in a systematic manner, giving due weightage to all related points.
- The technique for the purpose involves the undertaking of four major steps: (i) statement of the problem in a general way; (ii) understanding the nature of the problem; (iii) surveying the available literature (iv) developing the ideas through discussions; and (v) rephrasing the research problem into a working proposition.



11.7 FURTHER READING

1) Guthrie, G. (2010) Basic Research Methods: An Entry to Social Science Research, Sage Publications, New Delhi, India.

2) Kumar, R. (2011) Research Methodology, Sage Publications, New Delhi, India.

3) Nachmias, D., and Nachmias C. (2014) Research Methods in Social Sciences, St. Martin's Press, New York.



11.8 ANSWERS TO CHECK YOUR PROGRESS

Ans to Q No 1: A research problem is one which requires a researcher to find out the best solution for the given problem, i.e., to find out by

which course of action the objective can be attained optimally in the context of a given environment.

Ans to Q No 2: The four major steps in defining a research problem are: (i) statement of the problem in a general way; (ii) understanding the nature of the problem; (iii) surveying the available literature (iv) developing the ideas through discussions; and (v) rephrasing the research problem into a working proposition.

Ans to Q No 3: Review of literature helps a researcher to take new strides in the field for furtherance of knowledge i.e., s/he can move up starting from the existing premise. Studies on related problems are useful for indicating the type of difficulties that may be encountered in the present study as also the possible analytical shortcomings. At times such studies may also suggest useful and even new lines of approach to the present problem.



11.9 MODEL QUESTIONS

Answer the following questions in about 300-500 words.

Q 1: Discuss the process of identification of the research problem.

Q 2: Discuss the different factors that affect the selection of a research problem.

Q 3: Discuss the different techniques Involved in defining a research problem.

*** ***** ***

UNIT 12 : RESEARCH DESIGN

UNIT STRUCTURE

12.1 Learning objectives

12.2 Introduction

12.3 Meaning of Research Design

12.4 Components of Research Design

12.5 Need and Importance of Research Design

12.6 Features of a Good Research Design

12.7 Let Us Sum Up

12.8 Further Reading

12.9 Answers to check your progress

12.10 Model Questions

12.1 LEARNING OBJECTIVES

After going through this unit, you will be able to-

- understand the meaning of Research design
- discuss the need and importance of Research design
- identify the features of a good Research design.

12.2 INTRODUCTION

Research, as we all know, is the search for knowledge. It is the process through which one acquires knowledge about any natural or social phenomena. Sociologists too are continuously engaged in research relating to social phenomena. But for any kind of research to proceed in the right direction, there has to be a particular design or plan. Though the term design conveys several meanings, but in the context of research, it points to a plan, blueprint or outline of the research which specifies the objectives, methods of data collection and analysis, probable outcomes so on and so forth. The research design articulates what data is required, what methods are going to be used to collect and analyse this data, and how all of this is going to answer the research questions of a particular research. It is therefore

the statement of essential and fundamental elements of a study that provides basic guidelines of conducting the research. In the previous units we have learned about Qualitative and Quantitative Research. To conduct any of this type of research we need to have a research design. In this unit we will try to understand the meaning of research design, its need and importance in research, features of a good research design and the various types of research design.

12.3 MEANING OF RESEARCH DESIGN

Research Design is a very vital step in the research process. In simple words, the research design is an elaborated plan or model which throws light into how the entire research would be conducted. The Oxford Dictionary of Sociology defines 'Research Design as a strategic plan for a research programme, setting out the broad outline and key features of the work to be undertaken, including the methods of data collection and analysis to be employed and showing how the research strategy addresses the specific aims and objectives of the study, whether the research issues are theoretical or policy oriented'. Research design therefore stands for advance planning of the methods to be adopted for collecting the relevant data and techniques to be used in their analysis, keeping in view the objective of the research. Research design, infact, has a great bearing on the reliability of the results arrived at and as such constitutes the firm foundation of the entire edifice of the research work . Just as we need a ready architectural design or blueprint before the construction of a building so that we can have it as per our aspiration, similarly in case of research too, we need a research design to guide and direct us to achieve our results. The Research design will give a detail outline regarding the various aspects of the research from its objectives to the ways or methods in which data shall be collected and analysed. It would not only state the objectives of any particular research but also describe the way in which the research shall process to realize those objectives. It is therefore often acknowledged as the master plan for executing a research project Thus before the actual process of research or data collection begins, one must have the research design in hand to guide

him/her in collecting and analysing data.... In sociological research as well, framing a research design is the first and a very vital step. The research design is mostly about the following——

- What is the study about?
- Why is the study being made?
- Where will the study be carried out?
- What type of data is required?
- Where can the required data be found?
- What periods of time will the study include?
- What will be the sample design?
- What techniques of data collection will be used?
- How will the data be analyzed?
- In what style will the report be prepared?

12.4 COMPONENTS OF RESEARCH DESIGN

As discussed earlier, Research Design outlines how the research will be carried out. Therefore it addresses itself to certain key issues as mentioned below.

- Title of the study, '…should be brief, precise and project the scope in generalized terms.'
- Statement of the Problem, '…should be unambiguous, precise and the usage should be clear, simple and concise.'
- Review of Previous studies- '…should be a brief survey of a relevant literature in the subject concerned and presented subjectwise. Studies should be reviewed critically and from there further research is called for'
- Definition of Concept or Theoritical Principles involved, '…should be defined in general times and be linked with the study.'
- Coverage and Scope of the Study, '…should consider geographical, temporal and functional dimension.'
- Objectives of the study, '…should explain the main purpose precisely and in the form of question or an explanation to a particular issue.'

- Formulation of Hypothesis, '…should be empirical, conceptually clear, specific, close to things observable and related to the body of theory.'
- Methods of Investigation, '…depends upon the nature of study but researcher should define the methodology and techniques adopted.'
- Sampling design-'…definition of the universe or population, size of the sample and representatives of the sample should be defined.'
- Constructing of Schedule or Questionnaire-'…questions should be in a order or a form. Open ended questions are designed to permit a free response. The questions to be asked should have a direct bearing on the problem, avoiding the personal questions and multiple meaning.'
- Data Collection, '…depends upon the subject matter the unit of enquiry and the scope of the study.'
- Analysis of Data, '…to fulfil the objective or hypothesis the researcher should analyse the data and subject to the appropriate statistical analysis beside tabulation. Tabulation of results in a meaningful way is by itself a technique and an art. The data given in the tables must be in self-explanatory form
- Interpretation of Results-'…researcher should draw inference based on usual test for significance and relate with previous findings, for a wider field of generalization.
- Reporting and Findings: Reporting should be clear, specific, simple and related relating to the objective of the study. Researcher must report what has been discovered to ensure proper directions to other researchers in carrying out of similar researches.'

12.5 NEED AND IMPORTANCE OF RESEARCH DESIGN

From the above discussion we have already learned that Research design is the guide to research. It is, as we know by now, a broad plan of action which encompasses almost everything related to the way in which a particular research is to be carried out. Since the research is carried out as per the research design laid down in advance, hence it is very crucial in research. Let us now discuss in details the need for a research design:

We all know that if we travel to places which we have not visited earlier or

which we are not familiar with, we will need a map or a tour guide. Without a map or a tour guide, there is every possibility that we might get lost. The similar is the case with Research design. The significance of Research design lies in the fact that that it facilitates the conduct of the research by guiding the researcher on how he/she should proceed, how the researcher should collect data, which data are relevant, what should he/she do with the data so on and so forth. It gives direction to the research and helps the researcher to conduct an efficient research and gather as much data as required without wastage of time and money.. If a researcher conducts research without formulating a research design in advance, there is every possibility that he/she may go hay way. In this case, because the researcher won't have a clear plan regarding the way to proceed systematically, he/she might get confused regarding data collection and also might end up wasting time in collecting data not essential for that particular research. Thus he/she might have to spend a lot of time and money in the process which might just turn out to be a futile exercise. So it is very important to formulate a research design prior to conducting the research. As we all know that a building cannot be constructed without a blueprint or an architectural design, similarly there is also a need for a research design prior to the conducting of the research if it is to be an efficient research. Hence in absence of the research design there is every possibility that the researcher will face a lot of confusions and uncertainties and finally land up in a total mess.

Further, in absence of a research design, many researches may fail to serve the purpose for which they have been taken up. Research conducted without a research design may even lead to misleading conclusions. The Researcher should therefore very carefully prepare the Research Design in order to avoid futility of the research. He/she should prepare the Research design in such a way that the process of conducting research and collecting data becomes easier and at the same time scientific. Thoughtful designing of the research design would enable the researcher to avoid ambiguity.

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

Since in the research design the researcher would systematically organise his/her ideas and outline the entire strategy regarding how he/she is going to conduct the research, this would help him/her to identify and also rectify any errors that he/she might commit while conducting the research.. Thus the research design would not only facilitate the researcher to stick to the right path/direction of conducting research but also to correct the accidental mistakes that he/she might commit and which might mess up the entire research project. It is therefore utmost necessary to prepare an efficient and appropriate design before starting the process of data collection.

Another very important aspect of the research design is that it also helps others to get an idea about what that particular research is all about and how is it conducted. Going through the research design, one can give his/her reviews, comments and suggestions about the research and this is very beneficial for the researcher. The researcher can get comments and suggestions from his/her teachers, experts and also peers and incorporate them to enrich the study. He/she can also get to know about flaws and mistakes, if any, in his/her research and can correct it. In the absence the research design, it will be difficult for the critic to provide a comprehensive review of the proposed study.

From the above discussion it is understood that Research design is utmost necessary for conducting the research in an efficient, scientific and methodical way. Its importance is summarised in the following points:

- Research Design is like a map. It provides direction to the study. It helps the Researcher to proceed in the right direction and prevents the research from being merely a directionless and aimless empirical wandering.
- Formulation of Research design helps in minimising uncertainty, ambiguity, confusion and practical haphazard related to any research problem.
- It helps the researcher to collect and analyse data scientifically and systematically.
- It enables the researcher in identify and segregate relevant data's and facts from the irrelevant ones.

- In certain cases, the design acts as a tool to control the level of accuracy desired.
- It reduces inaccuracy increases the reliability of the research
- It prevents wastage of time and money.
- It provides a clear overview about the Research to others
- Gets rid of bias and marginal errors.



CHECK YOUR PROGRESS

Q 1: True or False :

a) Research design is formulated after conducting the research.

.....

.....

b) Research design minimises uncertainties and confusions

.....

.....

Q 2: Why is Research design similar to a map?

.....

.....

12.6 FEATURES OF A GOOD RESEARCH DESIGN

We know that the research design which provides direction to the researcher to carry out his/her research scientifically is very important for the research. Let us now try to understand the features of a good research design:

- Systematic, Scientific, Efficient, Economical and Flexible are all characteristics of a good research design.
- The design which gives clarity regarding what's and how's of research, which maximises reliability of data collected and analysed and minimises errors and biases is considered to be an appropriate and scientific research design.
- For a **good research design**, all the components of research should go together with each other in a coherent way. The theoretical and

conceptual framework must match with the research goals and purposes. In the same way, the data gathering method must fit with the research purposes, conceptual and theoretical framework and method of data analysis.

- Further, the design which gives the smallest experimental error is supposed to be the best design in many investigations.
- Similarly, the design which helps the researcher to gather maximum information or data and also 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 a 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 which may be quite appropriate for a particular study may not be the same for another study. One single design cannot serve the purpose of all types of research problems. But there are certain aspects that a good research design for a particular research should focus on and they are:

- The means of obtaining information
- The nature and skills of the researcher and his staff, if any
- The objective the problem to be studied
- The nature of the problem to be studied, and
- The availability of time and money for the research work



CHECK YOUR PROGRESS

Q 3: True or False

a) The research design which maximises bias is a good research design.

.....

.....

b) One single design cannot serve the purpose of all types of research problems.

.....

.....



12.7 LET US SUM UP

- The research design articulates what data is required, what methods are going to be used to collect and analyse this data, and how all of this is going to answer the research questions of a particular research.
- Research design therefore stands for advance planning of the methods to be adopted for collecting the relevant data and techniques to be used in their analysis, keeping in view the objective of the research.
- Just as we need a ready architectural design or blueprint before the construction of a building so that we can have it as per our aspiration, similarly in case of research too, we need a research design to guide and direct us to achieve our results.
- Research design enables the researcher to conduct the researcher systematically and scientifically.
- Research design facilitates efficient research and prevents wastage of time and money.
- Research design provides direction to the study. It helps in minimising uncertainty, ambiguity, confusion and practical haphazard related to any research problem.
- Absence of a research design, many researches may fail to serve the purpose for which they have been taken up. In fact lack of research design may even lead to misleading conclusions.
- A design is considered to be good if it can minimise bias and maximise the reliability of data collected and analysed.



12.8 FURTHER READING

1) Guthrie, Gerard, 2010, *Basic Research Methods: An Entry to Social Science Research*. New Delhi: Sage publication. India.

2) Kothari, C.R., 2004, *Research Methods and Techniques.* New Delhi:

New Age International Publisher.

3) Young, P.V., 2009, *Scientific Social Surveys and Research*. New Delhi: Prentice Hall.

http://www.yourarticlelibrary.com/marketing/research-design-introduction-contents-and-types/48714

opentextbc.ca/introductiontosociology/chapter/chapter2-sociological-research/

http://libguides.sun.ac.za/content.php?pid=426346&sid=3486790



12.9 ANSWERS TO CHECK YOUR PROGRESS

Ans to Q No 1: (a) False (b) True

Ans to Q No 2: Just as a map guides our way, and gives us direction similarly research design guides us regarding the way we should proceed in our research work

Ans to Q No 3: (a) False (b) True



12.10 MODEL QUESTIONS

A) Short Questions (Answer each question in about 150 words)

Q 1 : Define Research Design.

Q 2 : Explain the different components of a Research Design.

B) Long Questions (Answer each question in about 300-500 word)

Q 1 : Explain the need and Importance of Research design.

Q 2 : Discuss the different features of a good Research design.

*** ***** ***

UNIT 13 : TYPES OF RESEARCH DESIGN

UNIT STRUCTURE

13.1 Learning Objectives

13.2 Introduction

13.3 Exploratory Research Design

13.4 Descriptive of Research Design

13.5 Experimental Research Design

13.6 Diagnostic Research Design

13.7 Survey Methods

13.8 Case Study Methods

13.9 Difference between Quantitative and Qualitative Research Design

13.10 Let Us Sum Up

13.11 Further Reading

13.12 Answer to Check Your Progress

13.13 Model Questions

13.1 LEARNING OBJECTIVES

After going through this unit you, will be able to

- know about the meaning of exploratory research design and its importance in social research.
- understand the different components and concepts associated with case study methods.
- explain the diagnostic research design
- understand in detail the survey method of research designs
- Distinguish between quantitative and qualitative research design.

13.2 INTRODUCTION

The term ‘design’ means “drawing an outline” or planning or arranging

details. It is a process of making decisions before the situation arises in which the decision has to be carried out. 'Research Design' is planning a strategy of conducting research. It plans as to: what is to be observed, how it is to be observed, how to record observations, how to analyse/interpret observations, and how to generalize . Research design is, thus, a detailed plan of how the goals of research will be achieved.

A research design is the systematic and logical planning for action. It provides direction to research project. In precise, it may be called as a plan for collecting and analysing data in an efficient and relevant manner. It is a blue print and therefore at its best only tentative changes in the design are permitted and are dictated by considerations during the operations of the project. In other words, a research design is not a highly specific plan to be followed without deviation but rather a series of guide-posts to keep one headed in the right direction. Research design has been defined by different social scientists in different terms. In this unit, we will discuss in details about the various types of research design.

13.3 EXPLORATORY RESEARCH DESIGN

The main purpose of exploratory research is to gather information which will help in future for formulation of a precise research problem. On the basis of the collected facts the researcher may be able to formulate sound hypothesis for further research. It may also enable the researcher to get himself/herself acquainted with the phenomena which he/she expects to investigate at a later stage. The aim of an exploratory or formulative study may be clarification of concepts, establishing priorities for future research and collection of data about the actual conditions which affect an intended research. An exploratory research employs exploratory design.

Requirements of Exploratory Design

The essentials for exploratory or formulative design are:

- Review of pertinent literature: While proceeding in the path of research, the researcher has to take help from the work already done by his/her predecessors. By doing so he/she will not only save himself/herself

from the problem of trial and error but also minimize the expenditure of his/her energy.

- Experience Survey: Because of the complicated nature of social problems a researcher may not be in a position to collect all the required materials about a particular problem from one place. At times the researcher has to contact the persons who have earned enough experience to understand and analyse the social reactions.
- Analysis of Insight stimulating cases: Analysis of insight stimulating cases includes all those events, incidents and phenomenan that stimulate the researcher. Such cases invoke in the investigator the thinking required for the formulation of the hypotheses. In this regard the attitude of the investigator, intensity of the case study and integrative power of investigators appear to be very important.

13.4 DESCRIPTIVE RESEARCH DESIGN

The purpose of descriptive research design is to describe some event, situation, people, group or community or some phenomenon. Fundamentally it is a fact finding exercise which focuses on relatively far dimensions of a well defined entity, aiming at precise and systematic measurement of some dimensions of a phenomenon. Usually a descriptive design involves detailed numerical descriptions, such as distribution of a population of a community by age, sex, caste or education. The researcher may also take recourse to descriptive design for estimating the proportion of people in a particular geographical locality with respect to their specific view or attitudes.

However the procedure followed in descriptive design is broadly analogous, notwithstanding the differences evinced in their field, formulation of hypothesis, objectives, for treatment of the problem and in matters of field expansion.

13.5 EXPERIMENTAL METHOD

Experiment has been regarded as the sole basis of scientific method for fairly long time and it was on this basis that social sciences were denied the place of science. The use of this method has primarily been made in the physical sciences where the subject-matter can be given the laboratory treatment under controlled conditions. But this method which for a fairly long time remained the exclusive domain of the physical scientists is gradually finding favour with social sciences as well. Experiments now play an important part in researches in Psychology, Sociology and Anthropology, although social experiments are slightly different from the physical experiments as accounts of the limitations placed upon them by the nature of the subject matter with which the social sciences have to deal.

A social experiment has been variously defined by the writers. Jahoda and others, regard social experiment as a method of testing hypothesis. Greenwood suggests that, an experiment is the proof of a hypothesis which seeks to hook up two factors into a causal relationship through the study of contrasting situations which have been controlled on all factors except the one of interest the latter being either the hypothetical case or the hypothetical effect".

According to Festinger," the essence of an experiment may be described as observing the effect on a dependent variable by manipulation of independent variables."

The basic factor in case of experimental method is, thus the control over the subject of study and manipulation of an independent variable to study its effect upon the dependent variables. The control and manipulation of independent variable although fundamental to experimental method is nevertheless a rather loose term. It is not so much the control as such but the degree of control that one can exercise. It is because of this difference in the degree of control that there is wide variety of opinion regarding the social experiments.

Types of Experiments

Social experiments may be divided into various classes according

to the type of setting, nature and extent of control and the technique adopted for manipulating the variables. Greenwood has mentioned following five types of experiments.

- **Trial and Error Experiment**
  - This may be called as the earliest form of experiment. Under this method the researcher does not prepare a structured plan of study. He prepares a hypothesis and tries to test it on actual social conditions, wherever, they may be found. As the study is not rigidly planned his technique will undergo a constant modification in the light of experience. For example he may start his studies on a small group but later on finding that it was unsuitable may enlarge it or change it altogether.This type of experiment may hardly be called scientific and usually entails much fruitless labor and waste of time.
- **Natural Experiment**
  - It is also known as field experiment. This name has been given to this method because the experiment is conducted in the natural setting. It is different from the from the field study or field observation because in case of field experiment we try to manipulate the independent variable while in case of field study we simply observe the phenomena taking place without any kind of manipulation.
  - Natural experiment has been used for studying a wide variety of techniques and methods such as advertising technique, training methods, effect of group decision and group participation, effectiveness of political propaganda and the like. As an illustration we can cite the experiment carried on by Coch and French regarding the effect of participation by the employees in planning upon their resistance towards it. The hypothesis in this case was that greater the participation on the part of employees in planning to meet the changed condition, the less would be their resistance to the change. The experiment was conducted in a mill. Certain changes were introduced in the working of the

factory. In the cases of control group no participation of labor was allowed. The new changes were simply explained to the workers by the production manager. In case of the second group the representative of the workers were allowed to discuss things. In the third group total participation was allowed. In this way it was possible to study the effect of different type of participation and also varying degree of participation upon the resistance of the workers towards change. The hypothesis in this case was found to be correct as it was proved that in all cases where the workers were allowed to participate in deciding the change the resistance was almost negotiable.

- **Ex-post-facto Technique:** This technique is generally used to study the varying influence of two identical factors. Thus for example, suppose it is proposed to study the influence of Hindi and English education upon future adjustability in life, we shall take an equal number of students in each group keeping in mind that other variables like age, economic status, intelligence etc. should be matched in both cases. After, say, ten years we shall again try to locate those persons and see how they are placed in life. The difference between the two may be attributed to the nature of education received.
- **Laboratory Experiment:**
  - A laboratory experiment has been defined by Festinger' as one in which the investigator creates a situation with the exact conditions he wants to have and in which he controls some and manipulates other variables. The major difference between field experiment and laboratory experiment is that in field experiment the nature of experiment is studied in the natural setting with degree of control. A field experiment is studied in the natural setting with control over independent variables, but for a laboratory experiment the setting is artificially created. Thus, for example, if we make an experiment with the school boys in the school itself it would be field experiment, but if we create a small classroom and experiment upon it, it will be laboratory

experiment. In a laboratory experiment the phenomenan is artificially created and is naturally subject to greater control and manipulation.

13.6 DIAGONISTIC RESEARCH DESIGN

The diagonistic research design endeavours to find out the relationship between express causes and also suggests ways and means for the solution. Thus the diagonistic studies are concerned with discovering and testing whether certain variables are associated. Such studies may also aim at determining the frequency with which something occurs or the ways in which a phenomena is associated with some other factors.

Diagonistic studies are mostly motivated by hypothesis. A primary description of a problem serves the basis so as to relate the hypotheses with the source of the problem. Here only those data which form and corroborate the hypotheses are collected. With regards to the objective of diagonistic research design, it is based on such knowledge which can also be motivated or put into practice to find solution of the problem. Therefore it is obvious that the diagonistic design is concerned with both the cause as well as the treatment.

Diagonistic studies seek immediate to timely solution of causal elements. The researcher before going through other references, endeavours to remove and solve the factors and the causes responsible for giving rise to the problem.

The research design of diagonistic studies demands strict adherence to objectivity for elimination of any chances of personal bias or prejudice. Utmost care is taken while taking decisions regarding variables, nature of observation to be made in the field, the type of evidence to be collected and tools of data collection. Simultaneously the research economy should be carefully calculated. Any faulty decision in this regard will result in wastage of time, energy and money.

Usually the first step in such designing is accurate formulation of a research problem wherein the research objectives are precisely stated and principal areas of investigation are properly linked. Otherwise the investigator

will find it difficult to ensure the collection of required data in a systematic manner. Simultaneously, the clarification of concepts and the operational definition of the terms should be ensured so as to make them amenable to measurement.

At the next stage, certain decisions regarding collection of data are taken. In this regard, the researcher should always calculate in mind the advantages and disadvantages of the method to be employed and at the same time the nature of research problem, type of data needed, degree of desired accuracy etc. should be considered. Again, while collecting data, efforts should be made to maintain objectivity to the maximum possible extent.

In order to surmount problem of the financial constraints and paucity of time, a representative sample of the research universe should be drawn so as to gather relevant information. A wide range of sampling techniques are prevalent which must be made use of appropriately by the researchers.

At the stage of analysis of data, the researcher must take proper care in placing each item in the appropriate category, tabulating of data, applying statistical computation and so on.

13.7 SURVEY METHOD

The survey method is used to collect data of existing phenomena and to justify current practices or conditions or to make plans for improving them. The term 'Survey' has come from two words, 'Sur' and 'Vor', which mean ''to see'' a particular thing from a high place. In Natural Sciences it is used for measuring the things and in Social Sciences, it indicates the investigation of social problems, the technique of collection of data through interview, questionnaire and books etc. Many philosophers and scholars have defined the term in different ways. According to P.V.Young, " Social surveys are concerned with (i) the formulation of a constructive programme of social reform and (ii) amelioration of current conditions of a social pathological nature, which have definite geographic limits and definite social implications and significance, (iii) these conditions can be measured and compared with situations which can be accepted as models". It is claimed

that Young's definition of social survey is quite appropriate. Mark Abrams has defined social survey as a '' process by which quantitative facts are collected about the social aspect of a community's position and activities". Hence, from the above definitions it is clear that social survey in a broad sense provides the basis for theory construction, and the implications for social planning and reform.

The concept of social survey is not new. With a prolonged history, it is both ancient and persistent. The history of social survey may be dated back to 447-425 B.C. when Herodotus wrote his own account on the effects of climates upon the traditions, religions, beliefs and behavior of people along with the effects of natural disasters and war upon the social relations and economy. A German Scholar Earnest Engel, an economist and statistician by profession, contributed a lot to social survey by formulating an original theory based on the study of family budgets. In the United States Sorokin, Zimmerman and Trompton have contributed immensely to survey. Whereas Zimmerman and Trompton have applied Leplay's techniques in their study of Ozark regions and some other communities, Lind's 'Middle Town', Lloyd Warner's 'Yankee City' and Ralph Linton's 'Plain Valley' are categorized as geographical and cultural studies. Leplay's methods and techniques were also adopted by Charles Booth in England. By engaging the services of some young and energetic scholars like Beatrice Potter, Octavia Hill, Graham Balfour, Clara Collett, H. Llewellyn Smith and others, Booth conducted very remarkable surveys and presented a factual account of the matter. In all his works, Booth maintained objectivity and made empirical and scientific approach. Although H.C. Preston was inspired by Booth and made impressive studies of poor people, his book ''Bitter Cry of Outcast London''(1883) was more dramatic and emotional than realistic in character. A far more extensive study was made by B.S. Rowntree in the early 20th century.

Development of the survey movement in the United States may be dated back to 1890 with Jacob Ries' extensive work on the economic conditions of the slum dwellers in thickly populated towns. The year 1902 witnessed the publication of "Battle of the Slums". Thereafter "The

Commission for Rented Houses" was established in New York at the behest of President Roosevelt, in in order to collect relevant data regarding the conditions of people living in rental accomodations. Lincoln Steffen's survey of economic and political condition of seven industrial town of the states in1904 gave impetus to social survey movement of that country. In 1917 Mary Richmond published her classic book "Social Diagonsis", based on her social case work. She also tried to evaluate the significance of representative character of sampling. However Pacel Kellog and his associates conducted the most scientific survey in the first decade of the 20th century. Theirs was the most well organized and systematic survey which attempted to understand the nature and effects of urbanization and modernization upon society. Between 1927 and 1931 the Regional Planning commission of New York conducted a survey of the problems and difficulties of the people in rented houses. Philip Klein and his associates organized an extensive survey of the social service in Pittsburg and also a study of related problem. Among the private individuals who conducted Social Surveys in the United States, the names of Ann W. Shyne , Otto Pollak and Donald Young deserve mention

Stages of Social Survey Method

There are different stages for conducting a Social Survey. They are as follows:

- **Selection of Problem:** As social survey deals with social problems, selection of the problem is the first stage. Therefore unless and until the surveyors are not sure of the nature, character and clarity of the problem, they should not embark upon social survey. The social surveyor is not supposed to deal with every problem or any kind of problem. Rather as a careful observer the social surveyor must examine the clarity and definiteness of the problem he comes across in a community. Such a consideration must take into account the practical and utilitarian aspect which has its significance in the life of a community. Such a consideration must also take into account the practical and utilitarian aspect which has its significance in the life of a community.

- **Defining the purpose:** After the problem is chosen the surveyor attempts on defining the objective of the survey by making the purpose transparent so as to give direction to proceed, because in the absence of a crystal-clear definition of the aim the survey becomes directionless like a raddarless ship in the deep sea. The success of a survey and the capability of a surveyor is gauged only in relation to the attainment of goals. Hence when goals are not clearly defined the surveyor will grope in the dark. The goal may be generic or specific it may be for collection of census data or for knowing the living conditions of the working class specifically. In any case the goal must be well-defined.
- **Defining the Problem under investigation:** After having chosen the problem and having defined the goals, the surveyor endeavors to determine the scope, nature and character of the problem. For example, if the surveyor wishes to study the problem of drug addiction among the college youths he should define what he means by the term ‘’drug addiction’”. He should also clarify different types of drugs causing addiction and the quantity of drugs taken per day indicating addiction.
- **Preparation of a schedule:** The constituent factors involved in the problem and its various elements need to be analysed so as to present a constructive programme. This necessitates the formulation of a detailed and systematic list in order to gather the relevant information around as well as directly on the subject of study.
- **Constitution of a commission:** All the official surveys are undertaken after constituting a commission headed by a director and some members who determine the policy and oversee its execution. The commission for survey is empowered to collect various types of data relevant to the study. It has to access to classified material considered appropriate for the purpose of survey. If the situation warrants the commission may be divided and sub-divided into various committees and sub-committees to deal with specialized matters.
- **Setting limits:** Determination of scope is the hall -mark of any scientific study. Understanding a problem necessitates systematic investigation

into a limited and properly defined subject matter. Unless the scope is determined before the beginning of a survey, no phenomenan can be studied properly. Determination of scope takes into account the political and administrative division, different economic levels, biological bases, bases of social status, geographical base etc.

- **Setting time limit:** Not only cost estimate and trained investigators but also time estimate is essential for any successful survey. Whereas some problem need immediate solution , some others may require gradual transformation, some surveys are also required to be conducted within specific time intervals. The problems which are of topical importance undergo the process of gradual change. But some problems like displacement of population, resettlement etc. need to be tackled immediately.
- **Decision about the means of information:** Depending on the type of survey, different techniques of collection of information are employed. The preliminary data collected through primary survey are mostly obtained through questionnaire, schedule and even through direct interviewing of informants. On the contrary secondary data are obtained through secondary survey on the basis of official and non-official reports and also through published and unpublished materials of primary survey.
- **Determination of the units of survey:** In order to make the field workers free from any kind of dilemma and to ensure smooth collection of data, the unit of survey should be defined in definite and unambiguous terms. The units should be stable, harmonized, simple and feasible.
- **Refinement of Data:** As far as practicable, it must be determined before undertaking the actual survey as to what extent the materials collected can be refined. If it is a qualitative survey, the determination of accuracy and refinement should be directly proportional to the availability of manpower, finance and time. However despite the availability of the maximum amount of funds, time and man power, it is never possible to obtain cent per cent accuracy in qualitative survey.

- **Selection and training of field workers and researchers:** While making selection of the primary investigators, the nature and scope of social survey should always be borne in mind. Intellectual acumen should not only be the basis of selection. Along with that the selecter should also lay emphasis on the personal qualities like pleasant personality, a good sense of humour, tactfulness, ability to establish rapport with the informants and the like. After their selection, they will have to be properly trained so as to deal with various types of information in the field.
- **Preparedness for Cooperation of the Informants:** Unless the cooperation of the informants is obtained, there will the problem of non-response as well as want of accurate information . Therefore prior to the survey being actually undertaken, it is essential to cause preparedness and responsiveness among the informants so as to obtain accurate information by the fieldworkers. The informants must be assured that the information given by them will be kept strictly confidential.
- **Choice of Survey technique:** Determination of survey technique must be made before starting it, depending upon the scope and the nature of the survey. In case questionnaire or schedules are chosen, this must be prepared beforehand to avoid delay in surveying.
- **Conduct of Field Work:** After completion of the requirements as shown above, the field workers should get ready to go to the informants for collection of data. At this stage care must be taken to contact the informants keeping in view their convenience. The field workers should be properly trained to establish rapport with the respondents as to elicit accurate information. Supervision is also required in the field work.
- **Organization, Classification and Statistical Analysis of Data:** The Data collected by the field workers under the supervision of the supervisors are then organized, classified and statistically analysed so as to be interpreted.
- **Drawing Conclusion:** Conclusions are drawn from the findings based

on analysed data. It is always necessary take the help of competent and able researchers not only to draw conclusions, but also to elucidate the conclusions.

- **Graphical Representation:** The findings of a survey may be represented graphically so as to create lasting impression because of their attractiveness as well as ready intelligibility. Graphs have the capacity to bring to light the salient features of data at a glance without straining one's brain.

13.8 CASE STUDY METHOD

The case study method is a prominently used method of research. A case study can be defined as a holistic and in-depth study of an entity, a unit, person, event, etc. The case study research method enables the researcher to comprehend the subject in its entirety. The case study method is featured by its focus on bringing out the specificities and thereby developing a detailed understanding of the subject. It is perhaps this characteristic of the case study method that led Burgess to define the case study method as the 'social microscope'.

There are some common characteristics of the case study method. It facilitates the researcher to understand the uniqueness and particularities of the subject. This generally may not be possible in case of other research methods. While on the one hand, the case study research method enables the researcher to develop a general understanding of the subject, on the other hand, at the same time, it reflects the specificity of the subject. This further enables the researcher to carry out an intensive and detailed study. It is often seen that the utility of the case study method of research is compared to that of other methods, particularly the statistical method. It is seen that while the statistical method generally provides an overall view or idea about the study, the case study method on the other hand provides the entire details of the subject. In the domain of social science research, the entity of the 'individual' holds a significant place. In case of the statistical research method, the 'individual' becomes obsolete. On the contrary, in the case study method, the idea of the 'individual' becomes significant because

here the individual is considered as the whole and hence analyzed in its entirety. The case study method of research may employ several techniques while conducting a study. The techniques could be historical, descriptive, interview, etc, depending on the desired motive behind the study.

The case study method is also often analyzed in terms of the survey method of research. This is because the case study research is, often, also conducted as a supplementary method of study along with the survey method. In this context, the case study method can be further comprehended well by analyzing it in terms of the survey method. The case study, as mentioned above, caters to a detailed and intensive study, unlike the survey method which has a more broad based focus. This brings one to the aspect that while the survey method emphasizes on a plethora of units with the help of a selected sample, the case study method on the other hand emphasizes on a single unit. Further, while the findings of a study based on the survey method can be generalized, on the other hand, that of the case study method cannot be generalized. Again, the studies conducted on the basis of the survey method may be used to test hypothesis about large entities, social units, etc., those based on the case study method may be able to test hypothesis about specific entities.

13.9 DIFFERENCE BETWEEN QUANTITATIVE AND QUALITATIVE RESEARCH DESIGN

QUALITATIVE RESEARCH

Qualitative Research is a detailed, in-depth collection and analysis of quality data based on small samples, intended to provide insight and understanding.

Qualitative research addresses issues that grasp the subjective meaning from the perspective of the participants. It is a research strategy that usually emphasizes words rather than quantification in the collection and analysis of data. The primary aim of a qualitative research is to provide a complete, detailed description of the research topic. It is usually more exploratory in nature. The research situation is designed to be as open possible that enables the researcher to analyse the social situation. The

researcher serves as the primary data gathering instrument in qualitative research. Here, the researcher employs various data-gathering strategies, depending upon the thrust or approach of his/her research. Qualitative Research is primarily subjective in approach as it seeks to understand human behaviour and reasons that govern such behaviour. Thus, qualitative data is characterized by exhaustive and distinct attributes or categories that do not possess numerical magnitude. The researcher has the scope to use certain open techniques through which the person conducting the study collects information by observing phenomena and interacting with the respondents in its natural setting. Some of the most significant data-gathering strategies used in qualitative research are individual in-depth interviews, structured and unstructured interviews, focus group discussions, narratives, content or documentary analysis, ethnography or participant observation and archival research.

- **Ethnography/ Participant Observation:** The researcher involves oneself in a social setting for a considerable amount of time (sometimes months, years) to observe, experience and listen with the objective of having an in-depth understanding of a particular cultural group which is the subject of study.
- **Focus Group Discussions:** It is a kind of an interview with a group people on a specific topic or issue where each of the selected members discusses their views on the subject. It aims to reveal how the group participants view a particular issue that they might be facing.
- **Semi-structured and Unstructured Interviews:** Qualitative research makes use of semi-structured or unstructured interviews with open-ended questions. A respondent has the freedom to express his/ her views, feelings and experiences without any restriction, time or word limit. It helps the researcher to get an in-depth understanding of the research problem as the interviewee expresses hi/ her experiences and knowledge on the particular topic.
- **Content Analysis:** It is a technique to analyse documents and texts that seek to interpret information for the study. Content analysis is applicable to different forms of unstructured information such as

transcripts of interviews, open-ended questionnaires, case studies, archival materials etc. It seeks to bring out a pattern of information that can contribute to the research.

- **Case-study Method:** It is a research strategy that investigates a social phenomenon in its natural context. The "case" may be an individual, a community, a cultural group, an organization, an event or an action, existing in a specific time and place. The case study research design is useful for testing whether scientific theories and models actually work in the real world.

Advantages of Qualitative Research

- It provides liberty to the researcher to do an in depth study of the particular topic. The aim is to deliver a complete detailed description of a social reality.
- The semi-structured and unstructured interviews and participant observation methods employed in qualitative research helps in bringing about an unexplored domain of knowledge.
- The research design formulated for qualitative research may gradually evolve as the study advances and further explores the area of study.

Disadvantages of Qualitative Research

- Researcher tends to become subjectively immersed in the subject matter.
- Limited use of statistical data results in generation of non- statistical data in qualitative research which has less generalizeability.
- Qualitative research is more time consuming both in the processes of collection and analysis of data.

QUANTITATIVE RESEARCH

Quantitative Research techniques seek to quantify data and apply some form of statistical analysis.

Quantitative Research focuses more in counting and classifying features and constructing statistical models and figures to explain what is observed. Quantitative research techniques seek to quantify data and apply some form of statistical analysis. Quantitative data possesses the characteristic of numerical magnitude. Only measurable data are gathered

and analyzed in quantitative research and thus it is objective in its approach. It only seeks precise measurements and analysis of target concepts to answer his/her inquiry. Quantitative Research makes use of tools such as questionnaires, surveys, statistical measurements and other equipment to accumulate numerical or measurable data. Quantitative research is mostly based on hypotheses which are to be tested. A hypothesis formulates a relation between two or more variables or phenomenon that needs to be statistically tested in a quantitative research. Some of the important techniques of data collection and analysis in quantitative research are:

- **Hypothesis :** A hypothesis is a tentative assumption about relation between variables or an explanation of a research problem.
- *Close ended questionnaire:* Close-ended questionnaires gives options of specific answers one of which the respondent has to select. The multiple answers are already stated against each question. The close-ended questions are easier for quantitative analysis.
- **Structured Interviews:** Similar to the close-ended questionnaires, structured interviews consist of a set of questions of close-ended nature which the researcher asks the respondents. The respondents do not get much scope of elaboration of their views on a particular topic.
- **Experimental Research:** It attempts to establish a cause and effect relationship among the groups of participants that make up the independent variable of the study. In experimental research, the cause (the independent variable) is under the control of the researcher.
- **Use of Statistics:** Statistical methods are used extensively in pure science and social science disciplines. Quantitative research using statistical methods starts with the collection of data, based on the hypothesis or theory. Usually a big sample of data is collected – this would require verification and validation before the process of data analysis. Software packages such as SPSS (Statistical Package for Social Sciences) are used for this purpose.
- **Measures of Central Tendency:** It seeks to bring out an average for a distribution of values and has three different forms of average:- (a) Mean- here, all the values in the distribution are summed up and then

divided by the number of values. (b) Median- This is the mid-point in a distribution of values. (c) Mode- This is the value that occurs most frequently in a distribution.

- **Measures of Dispersion:** The amount of variation in a sample or a distribution of values is measured by using: (a) Range- it is the difference between the maximum and minimum value in a distribution of values. (b) Standard Deviation- The average amount of variation around the mean, it is calculated by taking the difference between each value in a distribution and the mean and then dividing the total of the differences by the number of values.
- **Correlation coefficient:** It is the measure of the relationship between two variables.
- **Multiple Regression:** It shows the degree to which multiple factors (independent variables) impact change in an outcome (dependent variable).
- **Chi-square test:** It is used to analyze data reported in categories, test is interested in seeing if two or more groups differ with respect to nominal variables. For example- the relationship between gender and educational qualification is statistically significant.
- **T-test:** *This* test used to see whether a difference between the means of two samples is significant.
- **ANOVA *(analysis of variance)* :** It is a technique which helps researchers discover if there is a significant difference between the means of more than two groups.
- Quantitative data may be represented in the form of tables, graphs and diagrams after analysis.

Advantages of Quantitative Research

- Data is in the form of numbers and statistical methods are used to procure objective results.
- Quantitative data has more generalizeablity is able to test hypotheses.
- Quantitative analysis requires lesser time compared to qualitative analysis and it can analyse result collected from huge sample size using statistical methods.

Disadvantages of Quantitative Research

- Use of close ended questionnaire and structured interviews in quantitative research may limit an indepth analysis of the research problem.
- This approach measures the responses of a large number of individuals to a limited set of questions. The analysis of relationships between variables creates a static view of social life.

Difference between Quantitative and Qualitative Research

Quantitative research generates numerical data or information that can be converted into numbers. Qualitative Research on the other hand generates non-numerical data. Quantitative research is based on the quantity or amount. It is applicable to phenomena that can be expressed in terms of quantity.

Qualitative research, on the other hand, is concerned with qualitative phenomenon that is relating to quality or kind. Qualitative research is especially important in the behavioural sciences where the aim is to discover the underlying motives of human behaviour. The following table shows some of the major differences between Quantitative and Qualitative research.

Table 13.1: Differences between Quantitative and Qualitative Research

Variables	**Quantitative Research**	**Qualitative Research**
Theory	As a starting point to be tested	As an end point to be developed
Data collection	Standardized	Open
Analysis of data	Statistical analysis of numerical data	Interpretative analysis, content analysis of interviews, discussions, observation, etc.
Knowledge generated	Establishes relationships, causation	Describes meaning, discovery
Report	Report statistical analysis. Basic element of analysis is numbers	Report rich narrative, interpretative. Basic element of analysis is words/ ideas, observation.



CHECK YOUR PROGRESS

Q.1: What is the main purpose of an exploratory or formulative design?

.....

.....

Q.2: What is descriptive research design?

.....

.....

Q.3: With what is the Diagnostic research design concerned?

.....

.....

Q.4: What is Experimental Research Design?

.....

.....

Q.5: What are the three components of Experimental Research Design?

.....

.....

Q.6: List one advantage and disadvantage of the Case Study Method.

.....

.....

Q.7: State the differences between Quantitative and Qualitative research.

.....

.....



13.10 LET US SUM UP

- Research design can broadly be classified into four types: Exploratory or Formulative Design, Descriptive Research Design, Diagonistic

Research Design and Experimental Research Design

- The main aim of an exploratory study is clarification of concepts, establishing priorities for future research and collection of data about the actual conditions which affect an intended research.
- The main purpose of descriptive design is to describe some event, situation, people, group or community or some phenomena.
- The diagonistic research design endeavours to find out the relationship between causes and suggest ways and means for the solution.
- The core of experimental research design lies in drawing inferences by observation of human relations under conditions of control.
- The survey method is used to collect data of existing phenomena and to justify current practices or conditions or to make plans for improving them.
- The case study method is a prominently used method of research. A case study can be defined as a holistic and in-depth study of an entity, a unit, person, event, etc. The case study research method enables the researcher to comprehend the subject in its entirety.



13.11 FURTHER READING

1) Ackoff, Russell L. 1961. *The Design of Social Research* Chicago: University of Chicago Press.

2) Ghosh, B.N. 1982. Scientific Methods and Social Research, New Delhi: Sterling Publishers Pvt.Ltd.



13.12 ANSWERS TO CHECK YOUR PROGRESS

Ans. to Q.No.1: The main purpose of exploratory or formulative design is to gather information which will help in future for formulation of a precise research problem.

Ans. to Q.No.2: A Descriptive Research Design is such a design which describes events, situation, people, group or community or some phenomena.

Ans. to Q.No.3: The Diagonistic Research design is concerned with the express characteristic and existing social problems.

Ans. to Q.No.4: Experimental Research design refers to systematic study of human relations by making the observation under conditions of control.

Ans. to Q.No.5: The three components of Experimental Research design are : comparison, manipulation and control.

Ans. to Q.No.6: Advantage: Research based on this method is easy to be conducted as it does not require any special condition or environment.

Disadvantage: The case study method is very time consuming.

Ans to Q. No 7 : Quantitative research generates numerical data or information that can be converted into numbers. Qualitative Research on the other hand generates non-numerical data. Quantitative research is based on the quantity or amount. It is applicable to phenomena that can be expressed in terms of quantity. Qualitative research, on the other hand, is concerned with qualitative phenomenon that is relating to quality or kind. Qualitative research is especially important in the behavioural sciences where the aim is to discover the underlying motives of human behaviour.



13.13 MODEL QUESTIONS

Q.1: What are the essentials of an exploratory or formulative study?

Q.2: Describe in detail about descriptive research design?

Q.3: Define experimental research design and describe its types?

Q.4: Discuss the case study method of research design.

Q.5: Distinguish between qualitative and quantitative research design.

*** ***** ***

UNIT 14 : THE RESEARCH REPORT

UNIT STRUCTURE

14.1 Learning Objectives

14.2 Introduction

14.3 Research Report: Meaning and Significance

14.4 Different Steps in writing the Report

14.5 Contents of a Research Report

14.6 Precautions in Writing Research Reports

14.7 Research Report Writing

14.8 Concept of Research Proposal and Research Abstract

14.9 Let Us Sum Up

14.10 Further Reading

14.11 Answers to check your progress

14.12 Model Questions

14.1 LEARNING OBJECTIVES

After going through this unit, you will be able to -

- understand the meaning of research report
- know about the contents of the research report
- identify the different steps in writing the research report
- understand the qualities of a good research report.

14.2 INTRODUCTION

Research Report, though the last, but is a very important part of research work. Any research work would remain futile if its findings are not documented. In the previous units, we have learned about the meaning of research, steps in conducting research, techniques and methods of data collection, analysis and presenting of data. In this unit we shall learn about and understand the meaning and importance of research report, contents of research report, steps involved in writing a research report and also qualities of a research report.

14.3 RESEARCH REPORT: MEANING SIGNIFICANCE

By now we have learned about the meaning of research, techniques of data collection and also the various methods of analysing data. But Research doesn't simply end with all these. A research work becomes meaningful only when the findings of the research are systematically written/ documented and presented. Analysis of data becomes futile and unrewarding if the research report is not written. Report writing is therefore very crucial to research. Though it is the last activity in a research work, yet it is of immense significance. The research work becomes meaningful and can be popularised only when the research findings are systematically assembled and written down with a valid conclusion. Research report writing is thus the culmination of the research investigation. It is a brief statement of most significant facts that are necessary for understanding the generalisations drawn by the researcher. Research report is prepared on the basis of the findings and the arguments drawn from it by the researcher with the intention of transmitting the information gathered to the wider and interested public. Efforts are made by the researcher to convey through his/her research study the entire result of the study in sufficient detail and so arranged as to enable each reader to comprehend the date and to determine for himself/herself the validity of the conclusions. Report writing not only enables mere dissipation of knowledge but also helps in further research in similar field.

14.4 DIFFERENT STEPS IN WRITING THE REPORT

Research reports are the final outcome of a lengthy process of collecting, analysing and chronologically arranging data and then finally transforming data into views and arguments. It therefore takes a long time to assemble the data or the findings and prepare a report. Several steps are involved in the process of preparing a report- a) logical analysis of the subject matter b) preparation of the final outline c) preparation of the rough draft d) rewriting and polishing e) preparation of the final bibliography f) writing the final draft. Let us now try to understand each one of them.

- ***Logical analysis of the subject matter:*** It is the first and a time consuming step in preparing the research report. There are two ways to develop the subject matter or the report: a) Logically and b) Chronologically. In the logical development, the researcher gives all his/her effort in establishing the logical connections and links between the findings of the study. He/she tries to relate one thing with the other by means of analysis. Secondly the researcher also tries to arrange the findings chronologically i.e based on a connection or sequence in time or occurrence. For example if a researcher is conducting research on some historical event, he shall first try to connect various incidences related to the event with one another and establish a logical connection among them. Secondly he/she shall try to arrange the incidences chronologically i.e which incident first took place and what fallowed after that. Arranging the data or findings of the study in the logical and chronological sequence gives clarity to the study and also depicts the analytical ability of the researcher.
- ***Preparation of the final outline:*** The step which follows the arrangement of the research findings in logical and chronological order is preparing the outline or the structure of the report. The preparation of outline of the report depends upon the judgement of the researcher. But it is important clearly explain the problem with which he/she was concerned, the method he/she applied to gather the findings, the analysis and conclusions and the bases for his conclusion. One the basis of the outline, the researcher would write down the report. Outline is therefore very important as it delineates the themes and sub themes to be covered. They are an aid to the logical organisation of the material and a reminder of the points to be stressed in the report. Without an outline, there is a possibility that the report may get jumbled up; it may not have a logical sequence or important points may be missed by the researcher.
- ***Preparation of the rough draft:*** This follows the logical analysis of the subject and the preparation of the final outline. This is perhaps the most time consuming of all the steps as the researcher now begins to

assemble his findings and thoughts from the research study and write down the report. What he/she has done; what methodology was adopted; what were the challenges; what are the findings, his/her views, arguments, generalizations and also suggestions are to be written down systematically. Since every relevant aspect of the research has to be incorporated, it takes much time of the researcher to put his/her thoughts into words

- ***Rewriting and polishing the rough draft:*** This is also one of the most difficult tasks in report writing and takes a longer time. In fact this step requires same time as that the writing of the rough draft. After the rough draft has been prepared, the researcher has to then carefully review it. While reviewing he/she would come across the weaknesses and flaws of the rough draft which shall have to be modified. Lot of rewriting, modification, and polishing of the rough draft takes place in this step to make the report flawless. Omissions and Errors of the rough draft including mistakes in grammar, spelling, sentence construction so on and so forth are rectified to bring the report to perfection.
- ***Preparing the bibliography****:* After the report has been brushed up; polished and final shape has been given to it, the next task of the researcher is to prepare the bibliography. The bibliography, which is generally appended to the research report, is a list of books in some way pertinent to the research which has been done. The researcher should carefully list the names of the books which he/she has consulted for the study in a proper instructed referencing style. This too is a very crucial stage because if the names of the books from which the researcher has taken ideas and materials from are left behind that he/she might have to face legal consequences of plagiarism.

14.5 CONTENTS OF THE REPORT

By now we know that once the research study is completed, the researcher has to write a report without doing which his/her entire research would be a meaningless effort. Now, in this section we shall learn about the

contents of the research report i.e. what does the research report contain. A reader who is reading the report must be communicated about the entire research study so that he/she can judge the adequacy of its methods and thus form an opinion of how seriously the findings are to be taken. Thus the report has to be written systematically for which a proper layout is necessary. The layout of the report means as to what the research report should contain. A comprehensive layout of the research report should contain a) Preliminary pages b) Main Body c) the Reference Section

PRELIMINARY SECTION: The preliminary section will vary according to the type of research one has undertaken. The preliminary section of the report usually comprise of the cover page indicating the Title, purpose of Research, Institutional Affiliation so on and so forth. The next page is usually the certificate page. Certificate page is usually found in dissertations and theses. It may not be required in other type of research reports. It is followed by acknowledgement page wherein the researcher offers gratitude to all those who have helped him/her to successfully complete the study. After this page there is the Content page which lists the details of contents in the report along with page numbers. The list of contents followed by list of tables and illustrations so that the decision-maker or anybody interested in reading the report can easily locate the required information in the report.

MAIN BODY: The main body of the research is the key component of the research report wherein the entire study and its findings are written down in a systematic manner. The main body is divided into the following subsections:

- ***Introduction:*** The purpose of the introduction is to introduce the research project to the readers. In the introduction, which is the opening section of the main body of the report, the researcher must state the reasons behind taking up that particular study.. The report should clearly explain the Background of the study i.e. what the study is about and why the study is important Depending on the particular piece of research, the review of the literature is introduced. This division will show what is known already and how our research will fill a gap in

knowledge or replicate earlier work. Further, the Objectives of the study, the Research Questions; Methods of data collection and analysis, and the theoretical Framework also needs to be described in the report. The researcher should give detailed elaboration of each of these so that the readers can easily comprehend about the study and judge its merit. Further, the scope of the study should be stated and the various limitations, under which the research project was completed, must also be narrated.

- ***Statement of the findings:*** This generally comprises the main body of the report, extending over several chapters.. After introducing the study in the introductory chapter, a detailed description of the findings of the study supported by data's, tables and figures should be given. Here the main task of the researcher is to convert the raw data derived from the study into meaningful statement and arguments. All relevant findings should be incorporated in the report and it should be written down systematically following a logical sequence. The analysis of the findings must also be clearly explained. The research report must state the findings of the study in a lucid, intelligible language.
- ***Implications of the results:*** Towards the end of the main text, the researcher should again put down the results of his researcher clearly and precisely. He/She should illustrate the implications that flow from the results of the study, for the general reader is interested in the implications for understanding the human behaviour. Such implications may have three aspects as stated below:
  - A statement of the inferences drawn from the present study which may be expected to apply in similar circumstances
  - The conditions of the present study which may limit the extent of legitimate generalizations of the inferences drawn from the study
  - The relevant questions that still remain unanswered or new questions raised by the study along with suggestions for the kind of research that would provide answers for them.
- ***Conclusion:*** The report remains incomplete without a conclusion.

Hence the researcher should conclude the study by summarising his/her entire study, its key findings also major arguments and generalisations. The researcher should also state the recommendations and suggestions if any in the conclusion. He/she also try to indicate the possibility and usefulness of future research in that particular field.

END MATTER: After completing the main text, the final task of the researcher, as we have discussed above, is to write down the bibliography. The researcher must carefully list down the books, articles, reports, government documents and all other secondary sources which he/she has referred to in the study. Apart from the bibliography, the researcher must also attach questionnaires, sample information, mathematical derivations, and the like ones. Index should invariably be given at the end of the report. The value of index lies in the fact that it works as a guide to the reader for the contents in the report. Photographs relating to the study, if any, should also be incorporated towards the end of the report with appropriate illustrations or captions.

14.6 PRECAUTIONS IN WRITING RESEARCH REPORTS

Apart from the contents of the report, there are several other technicalities involved in preparing the research report. There are definite and set rules which should be followed in the actual preparation of the research report or paper. The report must follow the instructions relating to Font size, Font Style, Line Spacing, Alignment, Margin, so on and so forth. Let us now try to understand some of the important technicalities that a researcher should confirm to while preparing the research report.

- **Size and Physical design:** The report should be on unruled, neat and legible paper. Equal margin must be kept on both sides of the paper. The format relating to font size, font style, line spacing, so on and so forth should be followed as per instruction.
- **Layout:** Keeping in view the objective and nature of the problem, the layout of the report should be thought of and decided and accordingly

adopted.

- **Quotations:** Quotations should be placed in quotation marks and double spaced, forming an immediate part of the text. But if a quotation is of a considerable length then it should be single – spaced and indented at least half an inch to the right of the normal text margin.
- **The Footnotes :** Footnotes are placed at the bottom of the page on which the reference or quotation which they identify or supplement ends. Footnotes are customarily separated from the textual material by a space of half an inch and a line about one and a half inches long. Footnotes should be numbered consecutively, usually beginning with 1 in each chapter separately. The number should be put slightly above the line, say at the end of a quotation. At the foot of the page, again, the footnote number should be used to correlate the reference in the text with its corresponding note at the bottom of the page, except in case of statistical tables and other numerical material, where symbols such as the asterisk (*) or the like one may be used to prevent confusion.
- Footnotes are always typed in single space though they are divided from one another by double space.
- Subsections must be clearly highlighted.
- In-text Citations should be properly given.
- Bibliography should be written in alphabetical order and must follow the instructed referencing style like APA etc..



CHECK YOUR PROGRESS

Q 1: What is meant by a research report?

.....

.....

Q 2: What is Bibliography?

.....

.....

Q 3: True or False:

a) Report writing is the first step in research

.....

.....

b) Acknowledgment page comprise of the lists of contents of the research report.

.....

.....

14.7 RESEARCH REPORT WRITING

The research report is indispensible if the research is to be meaningful. There is a great temptation to rush into writing because this gives a feeling of something accomplished and indeed it is true that certain sections of the report can sometimes be written in draft before the research is complete. But an illegible and confusing report gives a very bad impression about the researcher and his/her research study. Therefore the researcher should very cautiously write the research report incorporating all the relevant findings of his study in a systematic manner. He/she should use simple, uncomplicated language while writing the report. The report should be easily understandable and should represent the knowledge and wisdom of the researcher. Elaborations and illustrations must be provided wherever necessary. To be acknowledged and appreciated as a good research report, a report must have the following basic and essential qualities.

- **Make it easy to follow:** The report must be written in a way that it becomes easily understandable. Long paragraphs should be avoided and the paragraphs should focus only on one particular point or theme.. Trying to incorporate multiple points, ideas or themes in one particular paragraph may make it confusing and difficult to understand. The findings must be incorporated systematically and the paragraphs must follow a logical sequence.
- **Make it clear:** Clarity in writing is a quality of communication. A good research report must have clarity. The researcher must be able to clearly communicate his ideas and thoughts to the readers. Further

accuracy is also an essential quality in research report.

- **Use of Simple Sentences:** Sentences should be simple and short. It is better to use simple and precise sentences rather than long and confusing statements. The sentences should proceed methodically and should be effortlessly comprehensible.
- **Use of Non-technical language:** Avoid the use of technical terms; aim the report at the experience level of the common reader. . If it is necessary to use technical terms, a brief description or explanation should be included in the report.
- **Make it brief:** The report should be long enough to cover the objects of the study. Highlight major points by stressing the big issues and taking them up first.
- **Stress Practical actions:** Use analysis, specific examples, or compare drawn from experiences familiar to the reader. These show that the researcher has taken a realistic point of view.

14.8 CONCEPT OF RESEARCH PROPOSAL AND RESEARCH ABSTRACT

Research Proposal: After extensive review of literature on a proposed area/ issue for study, a research gap is recognized and a research problem is identified to be studied. Research proposal consists of the overall idea and plan for the study to be conducted. It includes introduction to the problem, rationale/ importance of the study, research questions, objectives of the study and methodology to be used.

Research Abstract: It is a brief summary of a research study conducted that includes introduction to the problem/ research topic, importance of the study, objectives and a brief explanation of the expected outcomes of the study. Research abstracts, within one or two paragraphs gives an overall idea about the study conducted regarding a particular research problem, without discussion of the results in detail.



14.9 LET US SUM UP

- A research work becomes meaningful only when the findings of the research are systematically written/documented and presented.
- Research report writing is the culmination of the research investigation. It is a brief statement of most significant facts that are necessary for understanding the generalisations drawn by the researcher.
- Research reports are the final outcome of a lengthy process of collecting, analysing and chronologically arranging data and then finally transforming data into views and arguments.
- Several steps are involved in the process of preparing a report- a) logical analysis of the subject matter b) preparation of the final outline c) preparation of the rough draft d) rewriting and polishing e) preparation of the final bibliography f) writing the final draft
- A comprehensive research report should contain a) Preliminary pages b) Main Body c) the Reference Section
- Apart from the contents of the report, there are several other technicalities involved in preparing the research report. The researcher must follow certain rules and instructions relating to the layout, citation style, and font style so on and so forth while preparing the research report.
- There are several qualities of a good research report. Some of the important qualities are: clarity, accuracy, preciseness, use of simple sentences, so on and so forth.



14.10 FURTHER READING

1) Guthrie, Gerard.(2010). *Basic Research Methods: An Entry to Social Science Research*. New Delhi: Sage publication. India.

2) Kothari, C. R. (2004). *Research Methodology – Methods and Techniques*, New Age International (P) Ltd, New Delhi.

14.11 ANSWERS TO CHECK YOUR PROGRESS

Ans to Q No 1: Research report writing is the culmination of the research investigation. It is a brief statement of most significant facts that are necessary for understanding the generalisations drawn by the researcher

Ans to Q No 2: Bibliography is the lists of books, articles, reports, government documents, and all other secondary sources that the researcher has consuted and reffered to in his/her study.

Ans to Q No 3: a) False b) False

14.12 MODEL QUESTIONS

A) Short Questions (Answer each question in about 150 words)

Q 1: Briefly explain the different technicalities involved in writing the research report.

Q 2 : Discuss the Qualities of a good research report.

B) Long Questions (Answer each question in about 300-500 words)

Q 1 : Discuss the different steps involved in writing a research report

Q 2 : Explain the main components of a research report

*** ***** ***



REFERENCES

1) Ahuja, Ram (2001): *Research Methods,* Rawat Publications, New Delhi.
2) Bernard, H. Russell (1994). Research methods in anthropology: qualitative and quantitative approaches (second edition). Walnut Creek, CA: AltaMira Press.
3) Creswell, J. W., *Qualitative Inquiry and Research Design: Choosing among Five Traditions*, Thousand Oaks, CA: Sage, 1998.
4) Chattopadhyay, Krichnadas & Sudarshana Sen (2013): *Rudiments of Social Research,* Levant Books, Kolkata.
5) DeWalt, Kathleen M. & DeWalt, Billie R. (1998). Participant observation. In H. Russell Bernard (Ed.), Handbook of methods in cultural anthropology (pp.259-300). Walnut Creek: AltaMira Press.
6) Fetterman, D. M., *Ethnography: Step by step*, 2nd edition. Newbury Park, CA: Sage, 1998.
7) Gupta, S.C. (2009). ' Fundamentals of Statistics, Himalayan Publishing House, Mumbai, Sixth Edition Reprint.
8) Gupta, S.P. (2010). 'Statistical Methods', Sultan Chand & Sons, New Delhi.
9) Kawulich, Barbara B. (2005, May). Participant Observation as a Data Collection Method [81 paragraphs]. Forum Qualitative Sozialforschung / Forum: Qualitative Social Research [On-line Journal], 6(2), Art. 43. Available at: http://www.qualitative-research.net/fqs-texte/2-05/05-2-43- e.htm [Date of access: June, 2017]
10) Kothari, C. R. (2004): *Research Methodology – Methods and Techniques*, New Age International (P) Ltd, New Delhi.
11) Marshall, Catherine & Rossman, Gretchen B. (1989). Designing qualitative research. Newbury Park, CA: Sage.
12) Ostrower, F. (1998). Nonparticipant Observation as an Introduction to Qualitative Research.
Teaching Sociology, 26 (1), 57-61. Retrieved from http://www.jstor.org/

stable/1318680 http://www3.ntu.edu.sg/home/patrick.williams/PDFs/Williams%20-%20Nonparticipant%20o bservation.pdf

13) Uwe Flick, Ernst von Kardorff, Ines Steinke *ed., A Companion to Qualitative Research,* London, Sage Publication, 2004.