IOSR Journal of Engineering (IOSRJEN) **www.iosrjen.org**
ISSN (e): 2250-3021, ISSN (p): 2278-8719
Vol. 11, Issue 4, April 2021, ||Series -I|| PP 09-16

Sampling and Hypothesis testing in Research Methodology

Anuja Lakhe[1], Shreya Chavan[2], Renuka Deshmukh[3], Jayshri Pagare[4]
M.tech, Department of Computer Science and Engineering, MGM university, Jawaharlal Nehru Engineering College, Aurangabad, India
M.tech, Department of Computer Science and Engineering, MGM university, Jawaharlal Nehru Engineering College, Aurangabad, India
M.tech, Department of Computer Science and Engineering, MGM university, Jawaharlal Nehru Engineering College, Aurangabad, India
Assistant Professor, Department of Computer Science and Engineering, MGM university, Jawaharlal Nehru Engineering College, Aurangabad, India
Received 26 March 2021; Accepted 09 April 2021

Abstract
In order to answer the research questions, it is doubtful that researcher should be able to collect data from all cases. Thus, there is a need to select a sample. This paper presents the steps to go through to conduct sampling. Furthermore, as there are different types of sampling methods, researcher needs to understand the differences to select the proper sampling method for the research. This paper also present different types of sampling techniques and methods. "Hypothesis testing is an integral and the most important part of the Research Methodology". In this paper, we introduce the overview of hypothesis and types of hypothesis. Researcher need the perfect method to reach towards the best conclusion so we also discuss the hypothesis testing steps, and various tools used for hypothesis testing so that researcher can take a perfect decision.
Keywords: Sampling, Sampling techniques, Hypothesis, Hypothesis testing, Hypothesis testing tools

I. INTRODUCTION

Research is defined as systematic gathering of data and information and its analysis for acquiring deep knowledge in any subject or field. In short, research methodology is the following various methods to search for knowledge through objective and systematic method of finding solution to a problem. Research methodology is the collection of steps should be followed to drawn proper conclusion about the concept or topic.
This is a set of systematic technique used in research. This simply means a guide to research and how it is conducted
This is a set of systematic technique used in research. This simply means a guide to research and how it is conducted
This is a set of systematic technique used in research. This simply means a guide to research and how it is conducted
This is a set of systematic technique used in research. This simply means a guide to research and how it is conducted
Researchers usually cannot make direct observations of every individual in the population. Thus, there is a need to select a sample. The entire set of cases from which researcher select sample and drawn conclusion about that sample is called sample and the entire cases which are used for study is called the population. Since, researchers neither have time nor the resources to analysis the entire population so they apply sampling technique to reduce the number of cases.

Hypothesis is nothing but the assumptions. Hypothesis is Greek word derived from the two words "hypo" it means low and "thesis" it means idea or theory expressed in logical way. Hypothesis testing is one of the important methods while drawing the conclusion regarding any problem statement. Every researcher must do the hypothesis testing for their research problem. There are two types of hypothesis which researchers are mostly used first is null hypothesis and second is alternative hypothesis, with the help of this and some hypothesis testing techniques researcher draw the conclusion**.**

II. STATISTICS

2.1. Descriptive Statistics:
It summarizes, show and describe the data in meaningful way. For example, in chart, graph, patterns etc. It is simply interpretation of data. Whenever there is lots of data for analysis then descriptive statistics is useful to visualize the distribution of data.
To describe the data there are two strategies

a. Measure of Spread
b. Measure of central tendency

2.2. Inferential Statistics

Inferential statistics takes data from a sample and makes inferences or draw conclusion about the larger population from which the sample was taken. It is help to make prediction from data. One important aspect is estimating parameters. This means taking a statistic from your sample data that is the sample mean and using it to say something about a population parameter that is the population mean.

For example, you might stand in a mall and ask a sample of 100 people if they like shopping. You could make a bar chart or pie chart of yes or no answers that would be descriptive statistics or you could use your research and inferential statistics to reason that around 75-80% of the population like shopping.

III. SAMPLING OVERVIEW

When we conduct research about the particular group of people, it's very difficult to collect data from every person in that group. Instead of that you can choose sample. The sample is the group of individuals who will participate in that research?

To draw valid conclusion from that sample, you have to carefully decide how you will select that.

It is process used in statistical analysis. Researcher collect the data for answering the research problem it is doubtful that researcher should be able to collect data from all cases, there is need to select sample. Sampling is a study of relationship existing between a population and sample drawn from the population. The main goal of sampling is to get as much information as possible about the whole population by examining only a part of it.

3.1. Population vs sample

First, you need to understand the difference between a population and sample.

- Population: is the entire group about which you want to draw conclusions
- Sample: is the specific group of individuals that you will collect data from.

3.2. Sampling Frame

The sampling frame is the list of individuals that the sample will be drawn from. Ideally, it should include the entire population.

Example, you are doing research on working conditions at College X. Your population is all 1000 student of the college your sampling frame is the college database which includes the lists of names and contact details of every student

3.3. Sample Size

The number of individuals you should include in your sample depends on various factors.

Example, we are doing research on student who get distinction in the final result so in this number of students included in your sample is the sample size.

IV. SAMPLING TECHNIQUES

Sampling can be used to make inference about the population. Sampling Techniques are divided in two types: [*Hamed Taherdoost et al.,2016*]

4.1. Probability Sampling:

It includes random selection of sample from the population, it allows you to make conclusion about the entire group. There are four types of probability sampling

4.1.1. Simple random sampling:

In Sample random sampling, every member of the population has the equal chance of selection. Sampling frame should include whole population while sampling. To conduct simple random sampling, we can use random number generators tools. **Example:** If we want to select 100 employees from the company then we select random 100 employee from the company database.

4.1.2. Systematic sampling

In Systematic sampling every member of population is listed like random sampling but instead of selecting randomly we choose member by following some pattern or interval. Example: All employees of the company are listed in alphabetical order and we select member whose name start and end with same letter.

4.1.3. Stratified random sampling

Stratified random sampling we divide population into subgroups called strata based on the properties (e.g., age, gender, height etc.). It allows you to draw more precise conclusion on the basis of sample which we made more correctly **Example**: If we want the number of male and female candidate working in the company then we can use stratified sampling to make samples based on their gender

4.1.4. Cluster sampling

Cluster sampling is type of sampling in which we divide population into multiple clusters for research. Each of these cluster contains member similar to each other, and these subgroups are called clusters. College it's very difficult for faculty to keep student academic, extra co-curriculum and health record so here faculty can make cluster of health record of every student.

4.2 Non-Probability Sampling:

It involves non-random selection based on convenience or other criteria and draw conclusion about the entire group.

4.2.1. Quota Sampling

Quota sampling is a non-random sampling technique in which participants are choose on the basis of predetermined characteristics. Total sample will have the same distribution of characteristics as the wider population. Example: You could divide population by state they live in, income or educational level.

4.2.2. Snowball Sampling

Snowball sampling is non random sampling method that uses few cases to help other cases to take part in the study, thereby increasing sample size. This approach is most applicable in small populations that are difficult to access due to their closed nature. Example: Secret societies and inaccessible professions

4.2.3. Convenience Sampling

Convenience sampling is selecting participants because they are often readily and easily available. Convenience sampling tends to be a favored sampling technique. It also helps to overcome many of the limitations associate with research. Example: using friends or family as a part of sample is easier than targeting unknown individuals.

4.2.4. Purposive or judgmental sampling

Purposive or judgmental sampling is a strategy in which particular settings persons or events are selected deliberately in order to provide important information that cannot be obtained from other choices. It is where the researcher includes cases or participants in the sample because they believe that they warrant inclusion. Consider a scenario where a panel decide to understand which are the factors which lead a person to select ethical hacking as a profession.

4.3. Sampling Process Steps:

Step 1: Clearly Define Target Population

This is the first step in sapling process in which clearly define target population. Population is commonly related to the number of people living in a particular country.

Step 2: Select Sampling Frame

A sampling frame is a list of actual cases from which sample will be drawn. The sampling frame must be representative of the population.

Step 3: Choose Sampling Technique

Sampling techniques can be divided into two types:

Probability or random sampling it involves techniques like Simple Random, Stratified Random, Cluster Sampling, Systematic Sampling, Multi Stage Sampling.

Non probability or non-random sampling it involves techniques like Quota Sampling, Snowball Sampling, Judgment Sampling, Convenience Sampling.

Step 4: Determine Sample Size

There are three factors used in determine the sample size

a. The margin of error
b. The confidence levels
c. The percentage of sample which chose a given answer to survey question.

Step 5: Collect Data

Once the target population, sampling technique, sampling frame and sample size have been established then the next step is to collect data.

Step 6: Assess Response Rate

Response case is the number of cases agreeing to take part in study. This are taken from original sample. Assess response rate help to reduce the like hood of sample bias.

V. HYPOTHESIS OVERVIEW

A hypothesis is the statement created by researchers when they contemplate upon the outcome of a research or experiment. It is an experimental statement or a prescribed statement of theory that shows how two or more variables are expected to relate to one another. Hypothesis is often considered as an important technique in research. Hypothesis is nothing but inference. It is presumption or guess based on the available information, which the researcher tries to prove through his study. It provides the analyst with a comparative statement that is

precisely testable in a research study. It prepares path to the researcher. It also provides structure to describe the result for the research study. A hypothesis will give a valid explanation that will be tested. A hypothesis may seem contradictory to the actual situation. It may prove to be true or false. Hypothesis need to be clear and definite and capable of being tested.

5.1. Types of statistical Hypothesis

There are two types of statistical hypothesis [*Sendil Mourougan, Dr. K. Sethuraman et al.,2017*]

5.1. Null Hypothesis

It is also known as the hypothesis of no difference. It is simply the hypothesis that sample observations result purely from chance. The null hypothesis, always states that the analysis has no effect, no change or no difference. According to the null hypothesis, the population mean after analysis is the same as it was before analysis. These are used when the researcher believes there is no relationship between two variables or one variable does not affect the other variable.

The null hypothesis represents an assumption that has been proposed, either because it is accepted to be true or because it is to be used as a basis for argument, but has not been proved. Has serious outcome if incorrect decision is made.

The null hypothesis is an entry point. The null hypothesis is an affirmation about a population parameter, such as the population mean, that is assumed to be true. We will test whether the value stated in the null hypothesis is likely to be true.

Null hypothesis is denoted by: H_0.

A null hypothesis is a statement that there is no actual relationship between variables. It may be given as, there is no difference between variables. H_0 states the opposite of what the researcher would expect or predict. The final conclusion of the researcher will either retain a null hypothesis or reject a null hypothesis in favor of an alternative hypothesis. Not rejecting H_0 does not really mean that H_0 is true. There might not be enough parameter against H_0. Once the null hypothesis has been stated, it is easy to construct the alternative hypothesis. It is essentially the statement that the null hypothesis is false.

5.2. Alternative Hypothesis

An alternative hypothesis is a hypothesis that directly opposite of the null hypothesis. It is done by describing that that the real value of a population parameter is less than, greater than, or not equal to the value stated in the null hypothesis.

The alternative hypothesis is a statement of what a hypothesis test is set up to establish.

Alternative Hypothesis is denoted by: H_1 or Ha.

It is opposite of Null Hypothesis. It is only reached if Ha is rejected. Frequently "alternative" is actual desired conclusion of the researcher.

An alternative hypothesis is a statement that suggests a potential outcome that the researcher may expect. (H_1 or Ha). It is established only when a null hypothesis is rejected. Often an alternative Hypothesis is the desired conclusion of the investigator.

The two types of alternative hypothesis:

5.1.1. Directional Hypothesis:

It specifies the direction of expected findings. Sometimes directional hypothesis is created to examine the relationship among variables rather than to compare groups.

5.1.2. Non-directional Hypothesis:

It is a type of alternative hypothesis in which no definite direction of the expected findings is specified. The researcher may not know what can be predicted from the past literature.

VI. HYPOTHESIS TESTING

Hypothesis testing is a technique which is used by statisticians to accept or reject statistical hypothesis. Here for hypothesis testing we use inferential statistics because it permits us to estimate behavior in samples to study more about the behavior in populations that are often too large or distant. The technique in which we select samples to learn more about characteristics or properties in a given population is also known as hypothesis testing. Hypothesis testing is a systematic approach to test assertion or interpretations about a sample or population

Hypothesis testing is also known significance testing. It is a method for testing an assertion or speculation about a parameter in a population, using data measured in a sample. The general objective of Hypothesis testing is to find out likelihood as a possible explanation for the result from research problem. [*Peter J. Veazie et al.,2015*]

6.1. Steps in hypothesis testing:

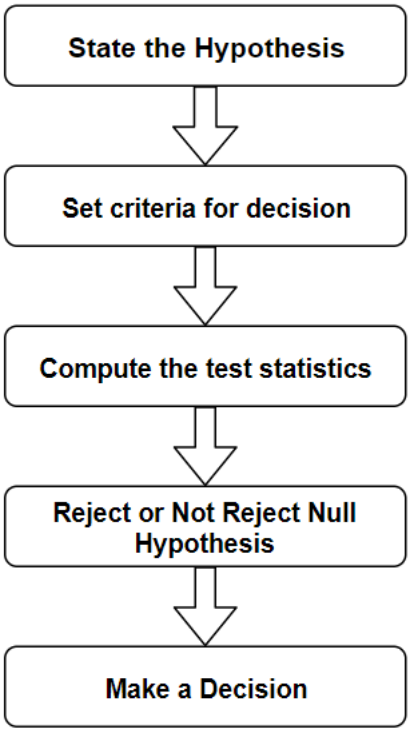

Figure 1: Steps in Hypothesis Testing

Step 1: State the hypothesis

This is the first step in hypothesis testing. This step involves the definition of null hypothesis and alternative hypothesis or stating null hypothesis and alternative hypothesis for the research study. These hypotheses must be mutually exclusive. It means that if one is true another must be false.

Step 2: Set criteria for decision

In this step researcher state the level of significance. It is criteria based on which decision is made with respect the value stated in null hypothesis. In hypothesis testing, we accumulate data to manifest that the null hypothesis is not true, it depends on the chance of selecting a sample mean from a population. Suppose researcher set 5% as the level of significance or probability in research studies. We reject the null hypothesis when the probability of getting sample, mean is less than 5% if the null hypothesis were true, then we conclude that the sample we selected is too unlikely. This step is used to decide whether the null hypothesis is rejected or accepted based on decision criteria.

Step 3: Compute the state statistics

Test statistics is nothing but the mathematical formula, it is used by researcher if the null hypothesis were true. To determine the probability of getting sample outcomes. To make a decision related the null hypothesis the result of the test statistic is used. There is various statistical test for example: -Z-test, t-test, f-test, chi-square test etc.

Step 4: Reject or not reject Null hypothesis

In this step we have to made decision regarding acceptance and rejection of null hypothesis. To decide acceptance or rejection we can use two method first is analysis of statistics state and second is computation of p-value.

Analysis of state statistics: -We consider the level of significance is 0.05 that is 5%. Compare this with state statistic test result if the result is greater than the corresponding values then result is significant and we reject the null hypothesis. If calculated result is less than the significance value then we accept null hypothesis.

P-value: - A p-value is nothing but the probability. When it is given that the value stated in the null hypothesis is true then we it used for getting a sample result. We use it for getting a sample outcome which is then compared with the level of significance. When the p-value is greater than 5% then we accept the null hypothesis. If the p-value is less than 5% then we reject the null hypothesis. [*Ashok Kumar et al.,2015*]

Step 5: -Make Decision
Final step where we finalize the decision based on above 4 steps. In the above steps we are deal with the sample data not on the entire population. The conclusion or result of entire population may be false or different than the sample result. So, we need to check whether there is error in result or not. Below table shows the types of error for making decision.

	Decision/Prediction		
		Accept Null Hypothesis	Reject Null Hypothesis
Actual	True	Correct 1-α	Type 1 Error α
	False	Type 2 Error β	Correct 1-β Power

Figure 2: Table for making Decision [*Frank Emmert-Streib and Matthias Dehmer et al.,2019*]

6.2. Decision:
Accept the Null Hypothesis
Our decision can be correct or incorrect when we decided to accept null hypothesis. If decision accept the null hypotheses is correct then actual result is true and decision is correct that is (1-α) .1-α is the probability of accepting null hypothesis that is actually true. If decision is incorrect then there is Type 2 Error (β). The Type 2 error is probability of accepting null hypothesis that is actually false.
Reject the Null Hypothesis
When we decided to accept null hypothesis, it can be correct or incorrect. If incorrect decision is rejecting the null hypothesis then it is Type 1 Error (α). Type I error is error when the null hypothesis is actually true but there is probability of rejecting a null hypothesis. Where α is level of significance. To reject a false null hypothesis is the correct decision. The null hypothesis is false when it is actually false, we are able to decide this with the help of probability. This decision is called the power of the decision-making process. [*KOLAWOLE, Yaoundé O et al 2017*]

6.2.2. Decision Rules:
Region of acceptance and Region of Rejection

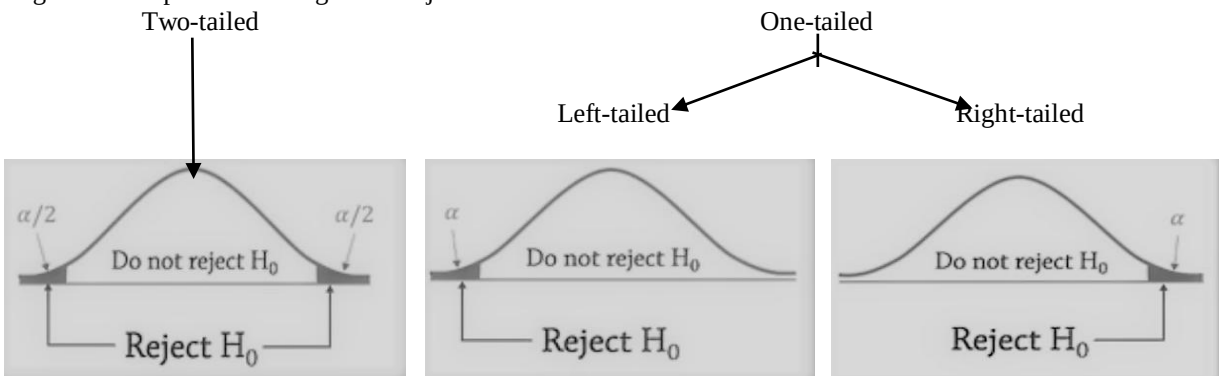

Figure 3: One-tailed and two tailed tests [*Sendil Mourougan, Dr. K. Sethuraman et al.,2017*]

VII. HYPOTHESIS TESTING TOOLS

Table 1: Hypothesis Testing Tools [*Joginder Kaur et al.*, 2015]

Tools	Information	Formula	
Z-test	It is used where sample size is large i.e., n>30.	Testing of significance for single mean $Z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$	Testing of significance for difference of means $z = \frac{\overline{X1} - \overline{X2}}{\sqrt{\sigma^2(\frac{1}{n1} + \frac{1}{n2})}}$

t-test	It was developed by William S. Gossett.	t-test for the mean of a random sample $t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$	t-test for difference of means of two samples $t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{(s^2(\frac{1}{n_1} + \frac{1}{n_2}))}}$	Paired t-test $t = \frac{\bar{d}}{s/\sqrt{n-1}}$
Chi Square Test	It was developed by Karl Pearson in 1900.It does not use mean, standard deviation. It is based on frequency	A test of goodness of fit $\chi^2 = \sum \frac{(Oij - Eij)^2}{Eij}$	A test of independence of attributes $\chi^2 = \sum \frac{(O-E)^2}{E}$	
F-test	It is used to test whether the two samples are from the same normal population with equal variance or from two normal populations with equal variances	$F = \frac{\sigma_1^2}{\sigma_2^2}$		
$\bar{x}$ -sample mean,μ-populatiuon mean, S -sample standard deviation ,σ-population standard deviation,(n,n1,n2)-degree of freedom,O-observed frequency,E-Expected frequency				

Example of Z-test

In year 2020 the mean age of all college student in city A has been 23. A random sample of 42 students, a mean age is 23.8. And population standard deviation of σ=2.4. Can we infer at α=0.05 the population mean age has changed?

Here, n=42, α=0.05, σ=2.4, $\bar{x}$ =23.8

Null Hypothesis, H_0: μ=23

Alternative Hypothesis, Ha: μ≠23

For α=0.05

Z=-1.96 or Z=1.96

$$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{23.8-23}{2.4/\sqrt{42}} = 2.16$$

Since Z=2.16>1.96, reject H_0

VIII. CONCLUSION:

This paper we describe the various sampling method and techniques with the various examples, using this we are able choose or select the perfect sample for the hypothesis testing. we also find the hypothesis testing steps. With the help of this we make decision regarding the research problem. We discuss the decision rules. We understood various tools used for testing the hypothesis, and explain it with some example.

REFERENCES:

[1]. Joginder Kaur," Techniques Used in Hypothesis Testing in Research Methodology – A Review", *International Journal of Science and Research*, Volume 4 Issue 5, May 2015.
[2]. KOLAWOLE, Yaoundé O," HYPOTHESES AND HYPOTHESIS TESTING", 30th June, 2017
[3]. Frank Emmert-Streib and Matthias Dehmer," Understanding Statistical Hypothesis Testing: The Logic of Statistical Inference", 12 August 2019.
[4]. M. Shafiqur Rahman," TEACHING HYPOTHESIS TEST IN AN EASY WAY", *International Journal of Business*, Economics and Management, December 2015.
[5]. Ashok Kumar," HYPOTHESIS TESTING IN MEDICAL RESEARCH: A KEY STATISTICALAPPLICATION", *Journal of Universal College of Medical Sciences* (2015) Vol.03 No.02 Issue 10.
[6]. Sendil Mourougan, Dr. K. Sethuraman," Hypothesis Development and Testing", *IOSR Journal of Business and Management*, Volume 19, Issue 5. Ver. I (May. 2017)
[7]. Peter J. Veazie," Understanding Statistical Testing", SAGE Open January-March 2015.
[8]. Hamed Taherdoost," Sampling Methods in Research Methodology; How to Choose a Sampling Technique for Research", *SSRN Electronic Journal* · January 2016.
[9]. PROBABILITY AND THE FOUNDATIONS OF INFERENTIAL STATISTICS.
[10]. http://southcampus.uok.edu.in/Files/Link/DownloadLink/RM%20U1%20P1.pdf

Anuja Lakhe, et. al. "Sampling and Hypothesis testing in Research Methodology." *IOSR Journal of Engineering (IOSRJEN)*, 11(04), 2021, pp. 09-16.