Department of AI & DS
CSE and CS&IT

COURSE NAME: PROBABILITY, STATISTICS AND QUEUING THEORY

COURSE CODE: 23MT2005

Topic

Testing of Hypothesis- Introduction

Session - 14

AIM OF THE SESSION

To familiarize students with the basic concept of testing of hypothesis

INSTRUCTIONAL OBJECTIVES

This Session is designed to:

1. Demonstrate the steps involved in testing of hypothesis
2. Describe role of statistical parameters in testing of hypothesis
3. Describe the difference between sample and population

LEARNING OUTCOMES

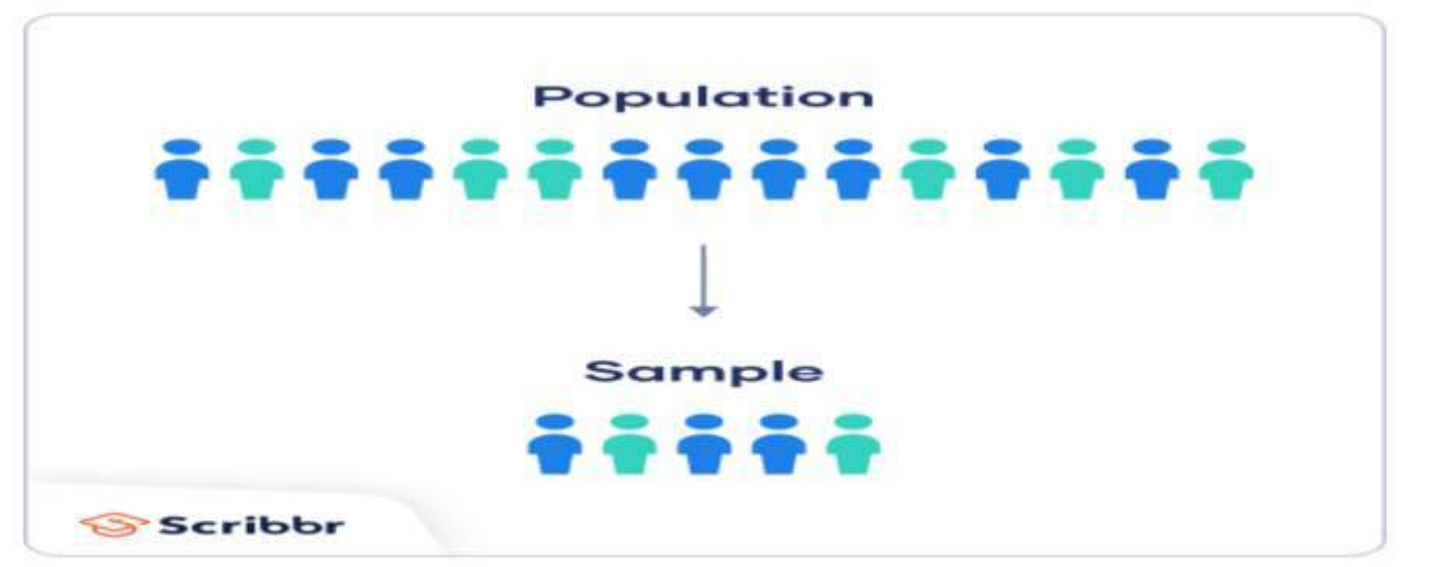
At the end of this session, you should be able to:

1. Define Null, alternative hypothesis.
2. State the difference between sample and population
3. Summarize the steps involved in testing of hypothesis

Population and Sample

A population is the entire group that you want to draw conclusions about.

A sample is the specific group that you will collect data from. The size of the sample is always less than the total size of the population.



Population and Sample

POPULATION (Big data or Huge or Massive data) A population is the aggregate of facts (Data). It is the set of all elements of interest for a specific Aim and Objectives. The process of conducting a survey to collect data for the entire population is called a census.

SAMPLE (Small data)

A sample is a subset of the population. The process of conducting a survey to collect data for a sample is called a sample survey.

In research, a population doesn't always refer to people. It can mean a group containing elements of anything you want to study, such as objects, events, organizations, countries, species, organisms, etc.

Difference between sample and population

The entire number of units of research interest is called population. In practical situations we are not taking all the population into consideration. We take some samples from the whole population and do the experiment for samples but give the conclusion to whole population.

A subset of the population under study.

- We distinguish between a sample mean and a population mean with the following symbols :
- $\bar{x}$: Sample mean
- μ: Population mean
- The formula for a sample mean is :

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

where n is the number of data points in the sample, the sample size.

- For a population, the formula is

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

- Where N is the size of the population.

Testing of Hypothesis

A statistical hypothesis is defined as a statement, which may or may not be true about the population parameter or bout the probability distribution of the parameter that we wish to validate on the basis of sample information.

For example: before the Loksabha or Vidhan sabha election process starts or just before the declaration of election results print media and electronic media conduct exit poll to predict the election result. In this process all voters are not included in the survey, only a portion of voter's i.e, sample is included to infer about the population. This is called inferential statistics.

In most of the scientific investigations a sample, a small portion of the population under investigation, is used for the study. On the basis of the information contained in the sample we try to draw conclusions about the population. This process is known as statistical inference.

Procedure for Testing of Hypothesis

1. **Null hypothesis:** Set up the Null hypothesis H_0
2. **Alternative hypothesis:** Set up alternative hypothesis H_1 , this will enables us to decide whether we have to use a single-tailed (right or left) test or two-tailed test.
3. **Level of significance:** Choose the appropriate level of significance (α) depending on the reliability of the estimates and permissible risk. This is to be decided before the sample is drawn, i.e., α is fixed in advance.
4. **Test statistic:** Compute the test statistic

$$Z = \frac{t - E(t)}{S.E.(t)}, \qquad \text{under } H_0$$

5. Conclusion: We compare the computed value of Z in step 4 with the significant value (tabulated value) z_α at the given level of significance, 'α'.

Testing of Hypothesis

Inferential statistics deals with drawing of conclusions about large group of individuals on the basis of observation of a few participants from among them or about the events which are yet to occur on the basis of past events. It provide tools to compute the probabilities of future behavior of the subjects.

There are two types of inferential procedures: (1) Estimation (2) Hypothesis testing.

Null Hypothesis: In hypothesis testing, a statistician or decision-maker should not be motivated by prospects of profit or loss resulting from the acceptance or rejection of the hypothesis. He should be completely impartial and should have no brief for any party or company nor should allow his personal views to influence the decision i.e., $H_0: \mu=\mu_0$

Alternative hypothesis: Any hypothesis which is complementary to the null hypothesis is called an alternative hypothesis, usually denoted by H_1. For example, if we want to test the null hypothesis that the population has a specified mean μ_0, i.e., $H_0: \mu=\mu_0$ then the alternative hypothesis should be:

(i) $H_1: \mu \neq \mu_0$ (i.e., $\mu > \mu_0$ or $\mu < \mu_0$) (ii) $H_1: \mu > \mu_0$ (iii) $H_1: \mu < \mu_0$

The alternative hypothesis in (i) is known as a two tailed alternative and the alternatives in (ii) and (iii) are known as right tailed and left tailed alternatives respectively.

The setting of alternative hypothesis is very important since it enables us to decide whether we have to use a single –tailed (right or left) or two tailed test.

In a particular problem, whether one-tailed or two-tailed test is to be applied depends entirely on the nature of the alternative hypothesis. If the alternative hypothesis is two tailed, we apply two tailed test and if alternative hypothesis is one tailed, we apply one tailed test.

One tailed and Two tailed tests

One-Tailed and Two tailed hypothesis: Depending upon the statement in alternative hypothesis (H_1), either a one-tail or two –tail test is chosen for knowing the statistical significance. A one –tail test is a directional test. It is formulated to find the significance of both the magnitude and the direction of the observed difference between two statistics. Thus, in two-tailed tests researcher is interested in testing whether one sample mean is significantly higher than the other sample mean.

Example: Researchers tested whether babies given the specially purified vitamin would walk earlier than babies in general.

the happiness example, the personality psychologists predicted the person who received $10 million would be happier than other people.

The researchers in these studies were not interested in the possibility that giving the specially purified vitamin would cause babies to start walking later or that people getting $10 million might become less happy.

IMPORTANT FACTS RELATED TO THE SESSION

The formula for the standard error of small sample mean score is as follows:

$$\text{S. E.}_{\text{m}} = \frac{\sigma}{\sqrt{n-1}}$$

The standard error of statistics has wide use in inferential statistics. It helps the experimenter or researcher in drawing concrete conclusions rather than abstract ones.

The various uses of standard error of the statistics are as under:

Various devices are used for determining the reliability of a sample taken from the large population. The reliability of the sample depends upon the reliability of the statistics, which is very easy to calculate.

The main focus of the standard error of statistics is to estimate the population parameters. No sampling device can ensure that the sample selected from a population may be representative. Thus the formula, of the standard error of statistics provides us the limits of the parameters, which may remain in an interval of the prefixed confidence interval.

IMPORTANT FACTS RELATED TO THE SESSION

Level of significance: the selection of the level of significance depends on the choice of the researcher. Generally level of significance is taken to be 5% or 1%.

The level of significance s($p<0.5$) is that probability of achance occurrence of observed results up to and below which the probability ‘p’ of the null hypothesis being correct is considered too low and the results of the experiment are considered significant ($p\leq$)..

SUMMARY

In this session, the basic concepts of testing of hypothesis have described

1. Define Null, alternative hypothesis.
2. State the difference between sample and population
3. Summarize the steps involved in testing of hypothesis

SELF-ASSESSMENT QUESTIONS

A hypothesis may be classified as

(a) Simple
(b) Composite
(c) Null
(d) All the above

Whether a test is one-sided or two-sided depends on

(a) Alternative hypothesis
(b) Composite hypothesis
(c) Null hypothesis
(d) Simple hypothesis

TERMINAL QUESTIONS

1. Describe basic steps involved in testing of hypothesis.

2. List out one tailed and two tailed tests.

3. Summarize Null and alternative hypothesis with suitable examples.

4. Importance of standard error of mean.

REFERENCES FOR FURTHER LEARNING OF THE SESSION

Reference Books:

1. William Feller, An Introduction to Probability Theory and Its Applications: Volime 1, Third Edition, 1968 by John Wiley & Sons, Inc.
2. Alex Tsun, Probability & Statistics with Applications to Computing (Available at: http://www.alextsun.com/files/Prob_Stat_for_CS_Book.pdf)
3. Richard A Johnson, Miller& Freund's Probability and statistics for Engineers, PHI, New Delhi, 11th Edition (2011).

Sites and Web links:

1. https://www.khanacademy.org/math/statistics-probability/significance-tests-one-sample/more-significance-testing-videos/v/small-sample-hypothesis-test

THANK YOU

Team – PSQT EVEN SEMESTER 2024-25