Statistical Hypothesis Testing: Tool for Drawing Meaningful Conclusions

Amit Shekhar
M Tech 4th (C.S.E.)
S.I.R.T. Indore M.P. India

Sachin Patel, PhD
Asso. Prof., C.S.E. Department S.I.R.T. Indore M.P. India

ABSTRACT
The hypothesis word is built with two different words first is Hypo and second is thesis. These two are combined to make the word Hypothesis. The meaning of first word is uncertain or question to the verification. The meaning of the second word is declaration about answer of a problem. A hypothesis should be specific and are testable so that the prediction can be done and we can expect it to happen in the study. To complete analysis of hypothesis there are three most important components first one is the variable used in the problem, second is size of the population, and last is what the relationship between the variables. The people who want to investigate can gowith carefully the truthful and theoretical elements that are related to the problem. Using this process we can decides the relevancy of truth. This process directs the efforts of the investigator into productive channels. This technique help and indicates what are the contents to look for is an investigation but how to obtain data. In this paper we presented some facts and study about the process of hypothesis. We also explain various term used in the hypothesis, like nature,importancecharacteristics, and limitation.

Keywords
Hypothesis, Variables, Population, Relationship, Facts, Statement

1. INTRODUCTION
A hypothesis is a kind of impermanent preparation to play substantial role in experimental research. It not only navigates research in a proper direction but also contributes in testing or suggesting correct theories and describing a social or lawful phenomenon. Hypotheses navigate a research and play important.It is the processes which do not care from the renounces are taken; it clearly states what the thing researcher wants to look for. It also describes probable relations between the concepts and between the variables which indicated therein. It suggested some of the reasonable descriptions about the facts. In fact, it directs the research. Without this process,it is not possible further step in experimental research and the researches are non-doctrinal research. This process helps in drawing some of the meaningful and fruitful conclusions which are supported by relevant and empirical data. This process provides a kind of guide to [9, 10] which work as

(i) collection of data in order to give answer the research problem;

(ii) Provides a kind of way to organized the data , data should be most competently and expressively

(iii) Provides a type of method which are used for data analysis and making decisions.

Once you tested a hypothesis it plays an important role.With the help it we give empirically statement some inferences can be drawn. With the help of this the initially posited relationship between the variables can be identify in the hypothesis. Consequently, we need to make some of the assumption, when we are going for empirically tested.Initially we assumed relationship between the concepts and variables, as the case may be, converts a verified fact. Once a this process is recognized, it concludes to be a suggestion.

2. FUNDAMENTAL CONCEPT
With the help of **Hypothesis Testing**, we develop a conclusion and we are to interpret a statement about the given population for which we are using sample data. This process evaluatestwo statements which are by nature mutually exclusive about a population.With the help of this we determine which statement is finest and supported by the sample size. Suppose we want to make some claims for the distribution of some data.Suppose we want to set results which are different from another results forsome machine learning problem, we apply and rely on statistical hypothesis tests [11,12].

The entire process of hypothesis is typically composed by

- Null Hypothesis: This is mathematically denoted by the symbol (H_0).This statement tells that all belongings remain equal. It is observed that no is relationship exist between variable.
- Alternative Hypothesis: This is mathematically denoted by the symbol (H_1). This statement tells the opposite of the first hypothesis. From this statement it is clear that there is some change present in the observed relationship and need to compare

$$H_0: \mu_A = \mu_B$$
$$H_0: \mu_A \neq \mu_B$$

This testing process determines the probability for a specific statement to evaluate true. It applied to evaluate twostatements which are mutually exclusive for some population and need to determine which statement best suitable and supported by the sample size.

3. FEATURES OF HYPOTHESIS TESTING
This process has a clear statement, that what we are going to discover and what we are intended to be investigated. It should we declared and specified before going to start research.We need to conduct it openly reporting the results clearly by the statement. This process allows andidentifies the objective of the research objectives, the key point for abstract and concepts involved in research work. It also shows review e relationship in both the problem statement and thereview work of literature. The following are some most important features of a hypothesis [13]

- This process has intangible nature.
- This process has an oral statement in the form of declarative nature.

- This process has the experimental referent and also indicates the uncertain association between the variables used in the problem.

4. SIGNIFICANCE OF HYPOTHESIS

Some of the important significances of hypothesis are[14,15]

1. **Its aim is give direction to the research:-:** Without this process, research has no direction and unfocussed research. This process serves as essential connection between concept and the research.
2. **It provides strong and precise objectives:** A well thought out set of hypothesis is that they place clear and specific goals before the research worker and provide researcher with a basis for selecting sample and research procedure to meet these goals.
3. **It provides a way to link facts and indication:** It provides important function to link facts and indication

5. CHARACTERISTICS FOR HYPOTHESIS

A hypothesis testing is best if and only if it must satisfy the following rules [2,4].

- This process never expressed in form question and we need to declare it to be empirically testable. It may be possible that the statement should be right or wrong.
- This process is always being specific and precise. It never creates a contradictory statement.
- This process always specifies relationship between variables and focus on variables which is to be established for the statement. It should describe one important fact only. We can form a hypothesis either in the form of expressive or interactive.
- This process provides assurances that existing tools and techniques are used effectively for the determination of confirmation and need to be stated as far as possible. The statement must be decaled in most simple terms which are easily understandable for all worried.
- This process can be explained the useful facts which gives to rise the need for clarification.

6. LIMITATIONS OF HYPOTHESIS

Hypothesis testing has several advantages and but there are also some limitations. Importantlimitations are as follows

1. The tests cannot be used in a mechanical style. It should keep in the mind that testing is not decision-making them; the tests can be useful and helps for decision-making process. Proper explanation of statistical confirmation is important to intelligent conclusions.
2. Test does not clarify the details as to why does the change exist, between the means for given two samples size. They simply indicate that the difference is due to variations of sample.There are other reasons also exist but the tests do not tell us which are the other reason and why causing these difference.
3. Outcomes of implication tests are based on likelihoods and cannot be communicated with full conviction. When a result of a test shows that a change is statistically important, then it only advises that the change is probably not due to unintended.
4. Statistical interpretations are built on the significance tests cannot be said to be entirely correct suggestions regarding the truth of the process. This is especially in case where small samples size and the likelihood of drawing erring inferences happen which is to be generally higher. For better dependability, the sample size used sufficiently inflamed.

7. LITERATURE SURVEY

In 2013 Md. Naseef-Ur-Rahman et al proposed **"Statistical Analysis based Hypothesis Testing Method in Biological Knowledge Discovery".** They showed that thequeries have potential role for discovering new relations and may open new area of search. They introduced a method based on text mining for answering the biological queries. These quires are expressed in terms of statistical computation. These queries help to researchers come up with new knowledge. It facilitates the user so that he or she can submit their query in natural linguistic. Proposed approach analyzed the hypothesis and measures p-value with respect to the existing literature. Hypothesis, it presents a network to give an integral overview of all the entities through which the entities might be related[1].

In 2014 John A. List et al proposed **"Multiple Hypothesis Testing in Experimental Economics"**.They provided a procedure which is based on bootstrap for testing null. They simultaneously used experiment andfor data random sampling.They used results of Romano and Wolf and showed that under weakassumptions procedure asymptotically controls the error rate.Probability of one or more false rejections and is asymptotically balanced the marginal probability.Proposed approach has much greater ability to detect truly false null hypotheses. They also showed how to exploit restrictions across nullhypotheses and how to further improve power[2].

In 2015 ValentinaCipriani proposed **"Ophthalmic statistics note 7: multiple hypotheses testing—to adjust or not to adjust".** Their main objective was to determine whether bevacizumabwas superior to standard NHS care. They performed single test to providedstrong evidence. They used evidence of a treatmenteffect for investigators and also to examine efficacyof visual function.They showed that a single question led tomany questions of interest and many testsof significance being proposed. They also showed that multiplicity may arise due to severaldifferent issues. Some of the issue are multiple outcomes, subgroups multiple time points, multiple questionsetc[3].

In 2016 John Concatoet al proposed **"P values: from suggestion to superstition"**They presented clinicians test to betterunderstand and gives evidence generated by research. They reviewedand define the p value, summarizes. They discussed corresponding issues including a conceptual approachof evaluating nullhypothesis. First they proposed the 1in20 chance of afalsepositive inference.The used A value ofp≤0.05, with other thresholds, and fou8nd that should not beemployed reflexively.They determined whether a clinicalresearch investigation is trustworthy from a scientificperspective. They also presented parallel issue with conceptualissues of validity and generalizability, quantitative. By the results they interpreted using a combinedassessment of strength of association, p values, CIs,and sample size[4].

In 2017 Chiang-Nan Chao1proposed **"An Examination on the Chinese Students' Rationales to receive their Higher Education in the U.S."**. They attemptedthis work to explore the bases why a lot of Chinese students select to education abroad and why theypreferredtheUnited States for r destination. This population is not very big but a vital factor of university life at several colleges.They have taken data of 380 students and filled with a questionnaire out of these 138 usable responses. Explicitly, the basis after Chinese students' foundations for attendance colleges in the U.Sis discovered. The outcomes direct that Chinese scholars are looking for education with a worldview, and opt to cessation fromthe Chinese scheme of knowledge. Although selecting to education in the U.S. is an theoretical attempt the

foundationsbehind that excellent is not merely for academics. A better sympathetic of the Chinese scholars' bases can benefitacademicians and college superintendents to improved target at this populace and attend them better[5].

In 2018 SanjoyDatta proposed "**Introduction to Hypothesis Testing".** They review the environment of hypothesis tests, as well as their clarifications. They include the position of sympathetic the fundamental wonder being verified. They described persistence of hypothesis testing is to select among two opposing hypotheses around the value of a populace limitation. They showed that hypothesis might assertion that the incomes of men and women are equivalent, though strength claims that men make more than women. B. The hypothesis essentially to be verified is typically given the symbol H0, and is commonly referred to as the null hypothesis. It is the hypothesis of no difference. It can be distinct as the theory which is below our thought and may have probable chances of refusal under the conventions which are true. It is a statistical hypothesis that states that there is no modification among a limitation and a exact value, or that there is no change between two restrictions [6].

In 2019 Jae H. Kim et al proposed **"Interval-Based Hypothesis Testing and ItsApplications to Economics and Finance".** They presented a transitory analysis of interval-based hypothesis testing, generally usedin bio-statistics, health science, and mindset, namely, tests for minimum-effect, correspondence,and non-inferiority. They presented the approaches in the environments of a one-sample t-test and a experiment forlinear boundaries in a recession. They also presented applications in challenging for market effectiveness, strength ofasset-pricing models, and perseverance of financial time series. They contend that, from the point of opinion offinances and finance, interval-based hypothesis testing affords more practical inferential outcomesthan those based on point-null hypothesis. They proposed that interval-based tests be normally employed in empirical research in commercial, as an substitute to point null hypothesis testing, particularly in thenew era of big data[7].

In 2020 David Delgado-Gómez et al proposed **"Improving the Teaching of Hypothesis Testing Usinga Divide-and-Conquer Strategy and Content ExposureControl in a Gamified Environment".**Theyproposed a new approach based on the divide-and-conquerprocedure to simplify its learning. The proposed approach is designed to consecutivelyevaluate and explain the different ideas involved in hypothesis testing. They ensurethat a newidea is not presented until the earlier one has been completely adapted. The proposed approachcontains several elements and implemented into anapplication through a modern game engine. Proposed approach was measured ina test in Statistics course within the IndustrialEngineering. The aimed the evaluating the acquiredknowledge, it was observed that students who used the developed application based on the proposedapproach [8].

8. METHODOLOGIES USED IN HYPOTHESIS

There are three methods of hypothesis testing are commonly used these are these are

Approach based on test Statistic

1. Decide test size α and calculate the critical value.
2. Calculate a experimental statistic
3. Discard H_0 if value of Test Statistic > Critical Value
4. Applicable interpretation.

Approach based on confidence Interval

1. Define H_0 and H_1
2. Define test size α or 1- α, and a hypothesized value
3. Construct the (1- α)100% confidence interval
4. Discard H_0 if a hypothesized value does not exist in Confidence Interval
5. Applicable interpretation.

Approach based on P-Value

1. Define H_0 and H_1
2. Decide test size α
3. Calculate a test statistic and p-value
4. Discard H_0 if p-value < α
5. Applicable interpretation

9. CONCLUSION

A hypothesis is the leading step in the course of scientific truth. In the pyramid of scientific information it is the lowermost on the scale. If experimental evidence can be found to confirm the hypothesis, it increases the status of a fact. A fact is the verified hypothesis. The assessment of hypothesis can best be understood by tracing their relationship to facts theories and laws. The experts build progressively a hierarchy of knowledge consisting hypotheses, theories and laws. In the proposed work we explain some important concepts of hypotheses. We explain some important terminology used in hypothesis, significance and limitation of hypothesis. We also expressed why hypothesis, is important for researchers.

6. REFERENCES

[1] Md. Naseef-Ur-RahmanChowdhury "Statistical Analysis based Hypothesis Testing Method in Biological Knowledge Discovery" International Journal on Computational Sciences & Applications (IJCSA) Vol.3, No.6, December 2013.

[2] John A. List _ Azeem M. Shaikh _ Yang Xu "Multiple Hypothesis Testing in Experimental Economics" John A. List Department of Economics, University of Chicago, 5757 S University Ave, Chicago, IL 60637, USA.

[3] ValentinaCipriani, Ana Quartilho "Ophthalmic statistics note 7: multiple hypothesis testing—to adjust or not to adjust" Cipriani V, et al. Br J Ophthalmol September 2015 Vol 99 No 9.

[4] John Concato, John A Hartigan "P values: from suggestion to superstition" Accepted 8 July 2016 Published Online First 3 August 2016 Copyright © 2016 American Federation for Medical Research.

[5] Chiang-Nan Chao1 "An Examination on the Chinese Students' Rationales to Receive theirHigher Education in the U.S." World Journal of Education Vol. 7, No. 3; 2017 .

[6] SanjoyDatta Concept of Testing of Hypothesis https://www.researchgate.net/publication/327885098SanjoyDatta on 26 September 2018.

[7] Jae H. Kim and Andrew P. Robinson "Interval-Based Hypothesis Testing and Its Applications to Economics and Finance" Econometrics 2019, 7, 21; www.mdpi.com/journal/econometrics.

[8] David Delgado-Gómez , Franks González-Landero "Improving the Teaching of Hypothesis Testing Using a Divide-and-Conquer Strategy and Content Exposure Control in a Gamified Environment" Mathematics 2020, 8, 2244; doi:10.3390/math8122244 www.mdpi.com/journal/mathematics.

[9] Bradley E. Alger Scientific Hypothesis-Testing StrengthensNeuroscience Research" July/August 2020, 7(4) ENEURO.0357-19.2020 1–10

[10] "Daniel Goldman "The Basics of Hypothesis Tests and Their Interpretations" August 2018 https://www.researchgate.net

[11] Frank EmmertStreib , and Matthias Dehmer "Understanding Statistical Hypothesis Testing: The Logic of Statistical Inference" 27 July 2019; Accepted: 9 August 2019; Published: 12 August 2019 Mach. Learn. Knowl. Extr. 2019.

[12] Bradley E. Alger "Hypothesis-Testing Improves the Predicted Reliability of Neuroscience" Research this version posted February 4, 2019.

[13] Jingyi Jessica Li and XinTong "Statistical Hypothesis Testing versus Machine Learning Binary Classification: Distinctions and Guidelines" Department of Data Sciences and Operations, Los Angeles, CA 90089, Correspondence: https://doi.org 2020.

[14] Bradley E. Alger "Hypothesis-Testing Improves the Predicted Reliability of Neuroscience Research" this version posted February 4, 2019. ; doi: https://doi.org/10.1101/537365 bioRxiv preprint.

[15] Jingyi Jessica Li and Xin Tong "Statistical Hypothesis Testing versus Machine Learning Binary Classification: Distinctions and Guidelines" 2Department of Data Sciences and Operations, Marshall School of Business, University of Southern California, Los Angeles, CA 90089, USA.

[16] Dr. Nilu Singh "Hypothesis Testing in Data Science**"** School of Computer Applications BabuBanarasi Das University Lucknow-UP.