

Demystifying Hypothesis Testing in Business and Social Sciences Studies

Dr. Isaac K. Naibei, PhD*

*School of Business & Economics, University of Kabianga
naibei2008@gmail.com

DOI: 10.29322/IJSRP.11.12.2021.p12047
<http://dx.doi.org/10.29322/IJSRP.11.12.2021.p12047>

Abstract- Hypothesis testing is probably one of the fundamental concepts in academic research especially where one wishes to prove a theory, logic or principle. Business and social research embeds several theories and established principles whose application in practice requires testing, and hypothesis testing comes in handy. Whereas the importance of hypothesis testing cannot be overemphasized, the procedure for hypothesis testing including setting up hypotheses by social scientists is not very clear. What is more, the conventional understanding of hypotheses testing may often lead to misleading interpretation. This purpose of this paper is to simplify the concepts is hypothesis testing. We use analysis of theoretical literature and simple illustrations to demonstrate use of hypothesis testing in business and social sciences studies. This paper concludes that a researcher wishing to test hypothesis should conceptualize a thesis or claim well in advance before collection of data. The claim can be drawn from theory or principles in various fields of study. We show that the only way to prove the claim under study is to prove that the counter claim is false. This is the rationale behind the setting of Null and Alternative hypothesis. We showed that it is the onus of the researcher to collect evidence to prove that the null hypothesis is false for us to conclude that our claim as stated in alternative hypothesis is true. Awareness of type I and II errors have also been highlighted as an important concern for researchers wishing to adopt hypothesis testing.

Index Terms- Null hypothesis, Alternative hypothesis, Level of significance, Type I error, Type II error

I. INTRODUCTION

Statistical hypotheses are a conjecture or speculation about a population parameter which may or may not be true. (Veazie, 2015) asserts that although statistical hypothesis testing is common in research, its conventional understanding sometimes leads to mistaken application and misinterpretation. They can also be viewed as testable propositions about the relationship between two or more variables. Hypotheses are predictions of some sort regarding the possible outcomes of the study. They indicate possible causes/reasons for the problem or assumptions or guesses drawn from the reading or from experience about the problem. Fisher's approach to statistical inference, emphasizes on the use data as evidence for or against the truth of a claim, provides a basis for hypothesis testing (Fisher, 1956).

Statistical hypothesis testing is commonly used when a researcher wishes to determine a substantive claim. If the truth or falsity of the substantive claim can be identified with the truth or falsity of a statistical hypothesis, then hypothesis testing can be used to inform judgments about the substantive claim (Veazie, 2015). Indeed, the logic behind hypothesis testing is that we have a thesis or claim but its truth or otherwise has not been established. It can only be established, when we have proof in form of statistical data. In order to clearly appreciate the process hypothesis testing, we start by considering some important concepts in hypothesis testing.

Concepts in hypothesis testing

a) Null and Alternative hypothesis (H_0)

Null hypothesis, symbolically expressed as H_0 , assumes a position of no relationship, no effect or no difference between a population parameter and a sample statistic. It is a proposition to be rejected and it is assumed to be true unless there is enough evidence to prove otherwise.

Alternative or Alternate hypothesis, symbolically expressed as H_a or H_1 , is a hypothesis that states that there is a difference, a relationship or effect between a population parameter and a sample statistic or two population parameters. It is a hypothesis that we tend to want to accept.

b) The level of significance

This is a very important concept in the context of hypothesis testing. It is always some percentage which should be chosen with great care, thought and reason. In case we take the significance level at 5% then this implies that H_0 will be rejected when the sampling result (observed evidence) has a less than 0.05 probability of occurring if H_0 is true. In other words, the 5% level of significance means that the researcher is willing to take as much as 5% risk of rejecting the null hypothesis when it happens to be true. Thus the significance level is the maximum value of the probability of rejecting H_0 when it is true and is usually determined in advance before testing the hypothesis.

c) Decision rule or test of hypothesis

After computing the sample statistic, we must decide whether to retain or reject the null hypothesis. Because we are observing a sample and not an entire population, it is possible that a conclusion may be wrong. There are four decision alternatives regarding the truth and falsity of the decision we make about a null hypothesis:

- (i) The decision to retain the null hypothesis could be correct.
- (ii) The decision to retain the null hypothesis could be incorrect.

- (iii) The decision to reject the null hypothesis could be correct.
- (iv) The decision to reject the null hypothesis could be incorrect.

d) Type I error

Type I error is the probability of rejecting a null hypothesis that is actually true. Researchers directly control for the probability of committing this type of error by stating an alpha (α) level. An alpha (α) level is the level of significance or criterion for a hypothesis test. It is the largest probability of committing a Type I error that we will allow and still decide to reject the null hypothesis.

Type I error (Alpha error) is the error of **incorrectly** rejecting the null hypothesis that is **actually true**. In other words, the error of *saying that a true statement is false*

e) Type II error

Type II error is also known as *Beta* error. It is the error of failing to reject a null hypothesis that is in fact, *false*. It is the error of saying that a false statement is true.

Table 1: Hypothesis Decisions and Errors

[1] Ho	is	[2] Decision	
actually		[3] Reject	[4] Fail to reject
[5] True		[6] Type I error	[7] OK
[8] False		[9] OK	[10] Type II Error

Illustration

To understand the types of errors and hypotheses, we take the defendant in court analogy. The criminal is usually considered innocent until proven guilty. The claim before the judge is that *‘the suspect committed murder’*. In this case the null hypothesis is:

H_0 : The suspect did not commit murder

The prosecution should prove, with evidence, beyond reasonable doubt that the suspect is indeed guilty of murder, otherwise the suspect will be freed (the null hypothesis stands as the prosecution has failed to reject it). The hypotheses which the researcher thinks should hold is the alternative hypothesis (in this case, *the suspect committed murder*). It can only be accepted when the researcher has collected data and statistically rejected the null hypothesis. Now look at the modified table of hypothesis and decision

Table 2: Type I and II errors in judgements

[11] Ho ‘the suspect did not commit murder	[12] Decision		
	[13] Reject	[14] Fail to reject	
[15] True (did not)	[16] Type I error(i)	I	[17] OK(ii)
[18] False (actually committed)	[19] OK(iii)	[20] Type II Error(iv)	

- (i) If we reject the hypothesis that H_0 : *The suspect did not commit murder*, when indeed the suspect was guilty of murder, the judge commits Type I Error (declaring that a true statement is false)
- (ii) If we fail to reject the hypothesis that H_0 : *The suspect did not commit murder*, when indeed the suspect committed murder, we make the right judgement (declaring that I true statement is true)
- (iii) If we reject the hypothesis that H_0 : *The suspect did not commit murder*, when indeed the suspect did

commit murder, we make the right judgement (declaring that a false statement is false)

- (iv) If we fail to reject the hypothesis that H_0 : *The suspect did not commit murder*, when indeed the suspect was guilty of murder, the judge commits Type II Error (declaring that a false statement is true)

Note: using the judges’ analogy, it is BETTER to acquit the guilty than to condemn the innocent. The burden is therefore put on the researcher to prove that the null hypothesis is false (to be rejected), so that the alternative hypothesis can be accepted otherwise, the null hypothesis stands.

As a researcher, you are really trying to prove H_1 but a principle of logic known as **fallacy of affirming the consequent** suggests that we can only prove “ H_1 ” indirectly by showing that H_0 is probably untrue (Nickerson, 2011).

Fallacy of affirming the consequent: A principle of logic that suggests that the only way to prove the alternative hypothesis is to demonstrate that the null hypothesis is untrue.

f) One-tailed tests

A one-tailed test is used when the researcher knows for certain that the outcome of an experiment is going to occur only in one direction or the researcher is only interested in one direction of the experiment. The null hypothesis uses directional expressions such as greater than, less than, older, younger, higher, lower, more, and so on which shows the direction the researcher is interested in. The alternate hypothesis will automatically take the opposite of the expression used in the null hypothesis. For example, if the null hypothesis used greater than, the alternate hypothesis will use less than.

Illustration:

A university wishes change its mode of study from face-to-face to blended learning. There is a claim the latter is very popular with the students and it is believed that at least 60% of the student population would prefer it. The management will launch the new mode of study only, if the claim is true.

H_1 : *At least 60% of the students will prefer blended learning to face-to-face mode*

H_0 : *Less than 60% of the students will prefer blended learning to face-to-face mode*

The above hypothesis is single tailed because we expect the results to be one sided or skewed to the right. In other words, we expect greater than (>) 60% of the sample to prefer blended learning. We will fail to reject the null hypothesis if less than 60% of the sample preferred blended learning.

g) Two-tailed

A two-tailed tests always use = and = in the/statistical hypotheses and are directionless in that the alternative hypothesis allows for either the greater than (>) or less than (<) possibility. For example, a machine that is used for packaging some product into containers of 500 grams is faulty. The engineers may not know whether it is overfilling or under filling the packages with the product. Therefore, a two-tailed test will be carried out.

The general model of two-tailed test is

$H_0: \mu = \mu_0$

$H_1: \mu \neq \mu_0$

Suppose a researcher wishes to establish the preference of a consumers to the Hybrid version of a particular model of car and the normal gasoline models. The researcher would set the hypothesis as follows;

H_0 : There is no significant difference in the consumer preference to the Hybrid and non-hybrid models ($H_0: \mu = \mu_0$)

H_1 : There is significant difference in the consumer preference to the Hybrid and non-hybrid models ($H_1: \mu \neq \mu_0$)

Steps in hypothesis testing

There are several steps involved in hypothesis testing. Different scholars have summarized the procedures into fewer steps, but the activities involved remain essentially similar.

Table 3: Steps of hypothesis testing

Step 1	State the hypothesis both the null and alternative
Step 2	Determine the appropriate statistical test
Step 3	State the level of significance
Step 4	Establish the decision rule
Step 5	Collect and analyze data
Step 6	Make statistical conclusion

We will use a simple illustration to explain the practical steps of testing hypothesis. Suppose a researcher wishes to establish the relationship between gender and average incomes of Kenyan working population. It is the argument of the researcher that there exists a significant income disparity between Kenyans across the gender.

Step 1: State the hypothesis both the null and alternative

To begin, we identify a hypothesis or claim that we feel should be tested. In this case, we might want to test the claim that there is significant income disparity between Kenyan male and female employees.

Null Hypothesis

H_0 : There is no significant difference between the incomes of males and females among Kenyan employees

Alternative hypothesis

H_1 : There is significant difference between the incomes of males and females among Kenyan employees

Step 2- Determine the appropriate statistical test

The next step would be to determine the appropriate statistical test such as Mean, proportion, variances and so on. For our illustration, it seems that the most appropriate test statistic is the **mean** of the incomes of Kenyan employees. Now, since we cannot possibly study the entire Kenyan employees, we take the **mean** of a carefully selected sample.

Step 3- State the level of significance

We now state the level of significance or the value of alpha or the Type I error rate that can be tolerated. The alpha level is the probability that the decision to reject a null hypothesis is incorrect. To minimize the probability of rejecting a null hypothesis that is indeed true, it is advisable to use lower alpha levels. In social

sciences an alpha level of 0.05 is recommended (this is 95% confidence level).

This means that there is a 5% chance that you will accept your alternative hypothesis when your null hypothesis is actually true. The smaller the significance level, the greater the burden of proof needed to reject the null hypothesis, or in other words, to support the alternative hypothesis.

Step 4- Establish the decision rule

If the p-value is less than the stated level of significance, the null hypothesis is rejected, otherwise we fail to reject the null hypothesis. The p-value describes the probability of obtaining a sample statistic as or more extreme by chance alone if your null hypothesis is true. This p-value is determined based on the result of your test statistic. Your conclusions about the hypothesis are based on your p-value and your significance level.

Example:

- P-value = 0.01 This will happen 1 in 100 times by pure chance if your null hypothesis is true. Not likely to happen strictly by chance.

Example:

- P-value = 0.75 This will happen 75 in 100 times by pure chance if your null hypothesis is true. Very likely to occur strictly by chance.

Generally, a p-value of <0.05 is set as threshold for business and social science studies.

Step 5- Collect and analyze data

This involves gathering data from the subjects that form the sample. For our illustration, we will gather data from a sample of Kenyan male and female employees. After collecting the data, we now analyze to compute the statistic. In our case we will compute the means and standard deviations of the income of male and female Kenyan employees.

Step 7- Make statistical conclusion

Using the decision rule set in step 4 and the computed value of the test statistic in step 5, a decision is made on whether to reject the null hypothesis or not to reject. We will, in our case compare the means of male and female monthly incomes and reach some conclusions. Let's assume we have computed and obtained the following values

Table 4: Means and Std deviations

	Male	Female
Mean monthly incomes (Ksh./pm)	$\bar{X}_1 = 11,300$	$\bar{X}_2 = 10,800$
Standard deviation	$S_1 = 600$	$S_2 = 500$
Sample size	$N_1 = 78$	$N_2 = 93$

We can see from the sample results that female respondents have less average incomes (Ksh. 10,800 pm) compared to (Ksh. 11,300 pm) for their male counterparts. At this point we may be tempted to reject the null hypothesis by refuting the claim that there is no significant difference between the average incomes of male and female Kenyan employees. However, it is too soon to reach that

conclusion. We need to establish whether that difference is significant or it just occurred by chance.

We compute the test statistic

Recall, we set our alpha at 0.05 and since our hypothesis is of the nature $H_0: \mu = \mu_0$

$H_1: \mu \neq \mu_0$, this is a two-tailed and the critical Z value corresponding 0.005 alpha is always ± 1.96 . We will therefore compute the Z value and compare with the critical value of ± 1.96 before reaching the conclusion.

Note

$$Z(\text{obtained}) = \frac{\bar{X}_1 - \bar{X}_2}{\delta \bar{X} - \bar{X}} \dots\dots\dots 1$$

$\bar{X}_1 - \bar{X}_2 =$ the difference in sample means

$\delta \bar{X} - \bar{X} =$ standards deviation of the difference of sample means given by;

$$\delta \bar{X} - \bar{X} = \sqrt{\left(\frac{S_1^2}{N_1-1} + \frac{S_2^2}{N_2-1}\right)} \dots\dots\dots 2$$

in our case $\bar{X} - \bar{X} = \sqrt{\left(\frac{600^2}{78-1} + \frac{500^2}{93-1}\right)} = 85.98$

Therefore, $Z(\text{obtained}) = \frac{11,300 - 10,800}{85.98} = 5.815$

We can now make decision

$Z(\text{obtained}) = 5.815$

$Z(\text{critical}) = \pm 1.96$

Note that the Z(obtained) lies beyond the critical zone of ± 1.96 , which leads us to the conclusion that the results (difference in incomes) are significant. We can now safely reject the null hypothesis which presumed no significant difference between the incomes of males and females. Given the direction of the difference (positive) we can also note that males average monthly incomes are significantly higher than females. The conclusion reached in step 7 is then applied in making recommendations for administrative or business decisions. For instance, we may recommend affirmative action to prioritize women in promotions to help move towards gender parity.

II. CHI SQUARE TEST OF INDEPENDENCE

The Chi-square test of independence is use to test independence of two categorical variables. In relation to chi-square independence refer to relationship between the variables and not the samples.

Two variables are independent if, the classification of case into a particular category of one variable has no effect on the probability that the case will fall into any particular category of second variable. In other words, two variables are independent if they are completely unrelated to each other.

Computation of Chi-Square

We can summarize two categorical variables within a two-way table, also called a $r \times c$ contingency table, where $r =$ number of rows, $c =$ number of columns. Our question of interest is "Are the two variables independent?" This question is set up using the following hypothesis statements:

Null Hypothesis

The two categorical variables are independent

Alternative Hypothesis

The two categorical variables are dependent

Chi-Square Test Statistic

$$\chi^2 = \sum (O-E)^2/E \dots\dots\dots (1)$$

where O represents the observed frequency. E is the expected frequency under the null hypothesis and computed by:

$$E = (\text{row total} \times \text{column total}) / \text{sample size}$$

We will compare the value of the test statistic to the critical value of χ^2 with degree of freedom $= (r - 1) / (c - 1)$, and reject the null hypothesis if $\chi^2 > \chi^2_{\alpha}$.

Illustration

Suppose we want to establish the question of whether gender is independent of education level. We collect a random sample of 395 people. Each person is asked to report the highest education obtained. Assume the survey results for the sample is as presented in the table:

Table 5: Distribution of respondents by gender and education

	Secondary	Bachelors	Masters	Ph.d.	Total
Female	60	54	46	41	201
Male	40	44	53	57	194
Total	100	98	99	98	395

We wish to answer the question 'is gender and education level dependent at 5% level of significance?' In other words, given the data collected above, is there a relationship between the gender of an individual and the level of education that they have obtained?

Now, we need to generate the expected values table using the formula $E = (\text{row total} \times \text{column total}) / \text{sample size}$

For instance, the expected count for row 1, column 1 = $(201 \times 100) / 395 = 50.886$

Similarly, expected count for r_2, c_1

$$= (201 \times 98) / 395 = 49.868$$

The same procedure will be used to fill all the cells in the matrix. The resultant table of expected counts is as follows

Table 6: Expected values

	Secondary	Bachelors	Masters	Ph.d.	Total
Female	50.886	49.868	50.377	49.868	201
Male	49.114	48.132	48.623	48.132	194
Total	100	98	99	98	395

The above expected frequencies table, tells us the frequencies we expect to get if the null hypotheses (H_0 gender and education level are independent) were true

We can now compute the Chi-square using the formula $\chi^2 = \sum (O-E)^2/E$

We can organize the information in the two tables above and determine chi-square as follows

<i>Observed Frequency (O)</i>	<i>Expected Frequency E</i>	<i>O-E</i>	<i>(O-E)²</i>	<i>(O-E)²/E</i>
60	50.886	9.114	83.064	1.632
54	49.868	4.132	17.070	0.342
46	50.377	-4.377	19.160	0.380
41	49.868	-8.868	78.648	1.577
40	49.114	-9.114	83.064	1.691
44	48.132	-4.132	17.070	0.355
53	48.623	4.377	19.160	0.394
57	48.132	8.868	78.648	1.634
N=395	N=395	00	χ^2 (Obtained)	8.006

Making Decision

The critical value of χ^2 with 3 degree of freedom is 7.815. Note that we read the critical value from distribution table of Chi-Square. Since 8.006 > 7.815, we reject the null hypothesis and

III. CONCLUSIONS

Here In this paper, we have demonstrated through easy illustrations the concepts of hypothesis testing. A researcher wishing to test hypothesis should conceptualize a thesis or claim well in advance before collection of data. The claim can be drawn from theory or principles in various fields of study. We show that the only way to prove the claim under study is to proof that the counter claim is false. This is the rationale behind the setting of Null and Alternative hypothesis. We showed that it is the onus of the researcher to collect evidence to prove that the null hypothesis is false for us to conclude that our claim as stated in alternative hypothesis is true. Another important element of hypothesis testing is inherent errors that can lead to wrong conclusion. The errors discussed in this paper are type one and type two errors. Type I error (Alpha error) is the error of **incorrectly** rejecting the null hypothesis that is **actually true**. In other words, the error of *saying that a true statement is false*. Type II error is the of failing to reject a null hypothesis that is in fact, *false*. It is the error of saying that a false statement is true (Kirui & Naibei, 2021).

conclude that the education level depends on gender at a 5% level of significance.

Note

Although we have demonstrated the procedure for obtaining the Chi-Square and making decision, this would help in understanding the logic behind Chi-Square but it is unlikely that you will use it in your study. This is because we now have sophisticated software which can compute it and report whether the difference is significant.

dispersion. Multiple regression analysis was carried out to determine relationship between independent and dependent variables in conformity with the prescription by a host of scholars (Faraway, 2002; Cohen, Cohen, Stephen, West, Leona & Aiken, 2003). Multiple regression analysis was preferred because it has ability to establish the existence of a relationship (positive or negative) between independent and dependent variables (Castillo, 2009). Further, regression would also establish the levels significance of the identified relationship. Pearson correlation measures the degree of association between variables. Attend conferences, workshops and symposiums on the same fields or on related counterparts.

ACKNOWLEDGMENT

The preferred spelling of the word “acknowledgment” in American English is without an “e” after the “g.” Use the singular heading even if you have many acknowledgments.

REFERENCES

- [1] Fisher, R. A. (1956). Statistical methods and scientific inference. New York: Hafner.
- [2] Kirui, J. K., & Naibei, I. K. (2021). Research Methodology for Social Sciences, a practical Approach. Nairobi: KLB.
- [3] Nickerson, R. S. (2011). Null-Hypothesis Significance Testing: Misconceptions. In M. Lovric, International Encyclopedia of Statistical Science. Springer, Berlin: Heidelberg.
- [4] Veazie, P. J. (2015, March). Understanding Statistical Testing. SAGE open, pp. 1-9.

AUTHORS

First Author –Dr, Isaac K. Naibei, University of Kabianga naibei2008@gmail.com

Correspondence Author – same as First Author