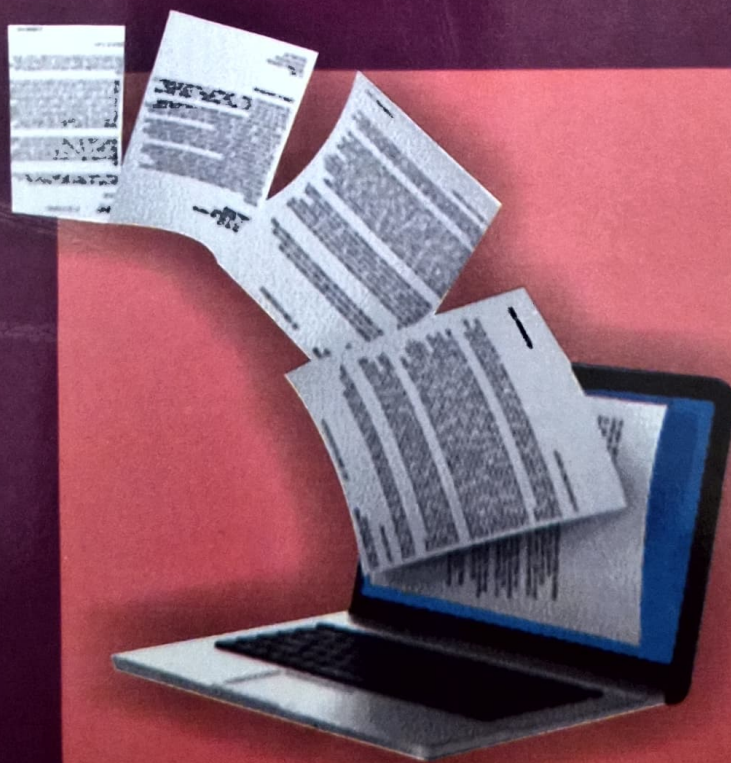




SOCIAL SCIENCES FOR BEGINNERS

VOL. III

Introductory Social Statistics and Computer Appreciation



Editors:
G.T. Ijaiya
U.A. Raheem
R.A. Bakare
F.M. Ajide

A Publication of the Faculty of Social Sciences, University of Ilorin, Ilorin Nigeria

TABLE OF CONTENTS

Chapter (Topics)	Page
Forward	
1. Introduction to Stata, Eviews and Spss Statistical Software Packages	2-13
2. Techniques of Data Presentation in Social Sciences	14-24
3. Introduction to Descriptive Statistics	25-35
4. Hypothesis Formulation and Testing	36-43
5. Computer Memory	44-51
6. Storage Devices and Parts	52-58
7. Introduction to Word Processing	59-69
8. Visualization and Presentation with Microsoft Power Point	70-81
9. Basic Knowledge in Microsoft Excel Spreadsheet	82-103
10. Microsoft Access	104-130
11. Challenges of Digital Literacy in Nigeria	131-141

Contributors Name and Departments

1.	Abdulhakeem A. Kilishi	Department of Economics
2.	Folorunsho M. Ajide	Department of Economics
3.	Toluwalope M. Agaja	Department of Geography and Environmental Management
	Issah, Moshood	Department of Sociology
4.	Adekeye D.S., Akangbe T.A., Akor	Department of Sociology
5.	Ameen, A.O	Department of Computer Science
6.	Ajewole David. O. and Agaja Toluwalope M.	Department of Geography and Environmental Management
7.	Bajeh, A O	Department Of Computer Science
8.	Mojeed H. A	Department of Computer Science,
9.	Mabayoje M. A.	Department of Computer Science
10.	Abdulraheem, Muyideen	Department of Computer Science
11.	K. O. Laro & U. A. Raheem	Department Of Geography And Environmental Management

Faculty of Social Sciences
 University Of Ilorin,
 Ilorin, Nigeria

CHAPTER FOUR

HYPOTHESIS FORMULATION AND TESTING

Adekeye D.S., Akangbe T.A., Akor

Department of Sociology
University of Ilorin, Ilorin

The Meaning of Hypothesis

Hypotheses are tentative explanations formulated as propositions from prior knowledge or as a result of some indications from the phenomenon under study. They are single tentative guesses, good hunches- assumed for use in devising theory or planning experiments intended to be given a direct experimental test when possible (Eric Rogers, 1966). Once these tentative statements are developed, a research work then verifies such hypothesis. They are usually formulated at the beginning of the research. A hypothesis is usually an explicit, clear, and testable tentative statement about possible scientific research outcomes. Such research is based on a population's specific characteristics, including the assumed differences among groups on identified variables or the correlation between variables (Ali & Bhaskar, 2016). Simply put, a hypothesis is a statement that can be tested by scientific research. A hypothesis is a conjectural statement of the relationship between two or more variables: a formal statement representing the expected relationship between an independent and dependent variable (Creswell, 1994).

A research hypothesis is a suggested explanation of a phenomenon or the prediction of an action's results. It is a reasoned proposal suggesting a possible correlation between or among a set of phenomena (Wikipedia, 2011:10) as cited in Obadan, 2012. It is an educated guess, formed as a statement that the researcher proposes to answer in the research question. A research question must be one which experimentation and observation in the natural world can provide needed information in order to be answerable. For example, a researcher may consider the incidence of armed robbery in Nigeria as a social rather than an economic issue. A hypothesis formulated based on this decision appears very broad, general and not capable of being tested satisfactorily. In order to get a good grip of this problem, however, this general hypothesis should be broken into several smaller problems, which are capable of empirical verification; hence the researcher may suggest that armed robbery in Nigeria is caused by poverty, poor parental upbringing, unemployment etc.

Other examples of such tentative statements are:

- Students who attend night classes will perform better than students who do not attend night classes.
- Students who sleep for up to 8 hours prior to an Examination will get higher scores than students who do not sleep for up to 8 hours.
- Drivers who drink before driving will more likely have an accident than those who do not drink before driving.

Before formulating a research hypothesis, the researcher should read about the topic of interest. From this reading, which may include research articles, books, and case studies, sufficient information may be gleaned, enabling the researcher to narrow or limit it and express it as a research question. The research question, when stated as one sentence, is the

research hypothesis. In the hypothesis, the researcher is predicting the relationship between variables. Through the disciplinary insight gained in the research process, the researcher will prove the hypothesis (Langkos, 2014). This is a process of discovery to create greater understandings or conclusions. It is not a strict proof as in logic or mathematics.

Characteristics of Hypotheses

1. They are not moral or ethical questions
2. They are the prediction of consequences
3. They are verbal statements in a declarative form
4. They can be tested, verifiable or falsifiable
5. They are considered valuable even if proven false

Features of a good hypothesis

1. It must state relationships between variables
2. It must contain clear implications for testing the stated relationships
3. It must include both independent and dependent variables
4. It must be testable by experiment, survey and any other scientifically sound technique

Types of Hypotheses

There are two types of hypotheses, they are:

1. Null Hypothesis (H_0)
2. Alternative Hypothesis (H_1)

The null hypothesis represents a theory that has been put forward, either because it is believed to be accurate or because it is to be used as a basis for argument, but has not been proved. The alternative hypothesis is a statement of what a hypothesis test is set up to establish. It is the opposite of the null hypothesis. Usually, the alternative hypothesis is the desired conclusion of the researcher. Using the examples of students and drivers given earlier, the null and alternative hypothesis are stated as:

• **Example 1**

"Students who attend night classes will perform better than students who do not attend night classes."

- H_0 There is no relationship between attending night classes and academic performance
- H_1 There is a relationship between attending night classes and academic performance

• **Example 2**

"Students who sleep for up to 8 hours prior to an Examination will get higher scores than students who do not sleep for up to 8 hours."

- H_0 There is no relationship between sleeping for 8 hours before an Examination and scoring higher marks
- H_1 There is a relationship between sleeping for 8 hours before an Examination and scoring higher marks

• Example 3

“Drivers who drink before driving will more likely have an accident than those who do not drink before driving.”

H_0 There is no relationship between drinking before driving and the likelihood of having an accident.

H_1 There is a relationship between drinking before driving and the likelihood of having an accident.

When formulating a hypothesis, the steps to follow is determined by the kind of research study being conducted. Familiarization with the topic is essential before making the final decision. The researcher needs to be confident that the topic is researchable; that is, there are sufficient published materials on the topic. After collecting as many observations as possible about the research problem or topic, an evaluation is made about these observations. This evaluation is done in order to look for possible causes of the research problem. Once a list of possible explanations has been generated, the hypothesis can now be stated and ways of confirming or refuting each possible explanation through experimentation is done (Castillo, 2013).

Hypothesis Testing

Hypothesis testing refers to how the results of an experiment or a survey are tested to determine if there are any meaningful results. Here, the researcher tests whether the results are valid by figuring out the odds that such results have happened by chance. Hypothesis testing also called significance testing is a method for testing a claim or hypothesis about a parameter in a population, using data measured in a sample (1). In hypothesis testing, the study is conducted to test whether the null hypothesis is likely to be true. The steps involved in hypothesis testing are as follows:

1. Specifying the Null Hypothesis (H_0) and the Alternative hypothesis (H_1)

A Null hypothesis is a statement of no effect, relationship or difference between two or more groups. Generally, a researcher is usually interested in disproving the null hypothesis. Examples of Null hypotheses are:

- i. There is no difference in armed robbery rates in Ghana and Nigeria.
- ii. There is no relationship between poverty and armed robbery in developing countries.

An Alternative hypothesis is a statement that has effect or difference. It is a statement that clearly contradicts a null hypothesis by stating that the true value of a sample parameter is not the same as, less than or greater than value stated in the null hypothesis. Usually this is the hypothesis that the researcher is interested in proving. Examples of Alternative hypotheses are:

- i. There is a difference in armed robbery rates in Ghana and Nigeria
- iii. There is a relationship between poverty and armed robbery in developing countries.

2. Setting the criteria for a decision. (Significance Level)

In setting the criteria for a decision the significance level for the test must be stated.

The level of significance is usually set at 0.05. Meaning there is a 5% chance of accepting the alternative hypothesis when the null hypothesis is indeed true. It is important to note that the smaller the significance level, the greater the burden of proof that is needed to reject the null hypothesis and support the alternative hypothesis.

3. Calculating the Test Statistic and Corresponding P-value

The P-value is a probability measure of finding the observed, or more extreme results when the null hypothesis of a given statistical test is true. For instance where the P-value is 0.02, this means the effect will happen 2 in 100 times based solely on chance if the null hypothesis is true. That is, not likely to happen by chance. However, where the P-value is 0.65, this means it will happen 65 in 100 times by chance if the null hypothesis is true, that is, it is very likely to occur by chance.

The sample size of the research directly impacts the P-value, because small sample sizes produce high P-values and larger sample sizes produce small P-values. This can occur even when the differences between groups are not meaningful

4. Drawing a Conclusion

Drawing conclusions to any hypothesis test is dependent on the statistical significance of the result. When the P-value is less the significance level, the null hypothesis is rejected in favour of the alternative hypothesis. The hypothesis test reveals that the result is statistically significant. On the other hand when the P-value is greater than the significance level, the null hypothesis is not rejected, and the test reveals that the result is not statistically significant. When there are no evidences against the null hypothesis, then the null hypothesis cannot be rejected, but when the test produces evidences against the null hypothesis, it can be rejected. When the p value is less than 0.05, a level of significance is reached and the conclusion is stated in favour of the null or alternative hypothesis based on the P-value and significance level.

Errors in Hypotheses

After drawing the conclusion from a test, there are four likely outcomes that may emerge. This is because the study is only observing a sample of the population and not its entirety. The conclusion of retaining the null hypothesis could be correct; the conclusion of retaining the null hypothesis could be incorrect; the conclusion to reject the null hypothesis could be correct, and the conclusion to reject the null hypothesis could be incorrect. This may further leads to errors in the hypotheses. There are two types of mistakes that may occur while testing hypotheses. They are Type I and Type II errors.

Type I errors occurs when the null hypothesis is wrongly rejected. In order to minimize the likelihood of type I error, an assumption is made at the beginning of a hypothesis test that the null hypothesis is true. Once this assumption is made, Type I error is controlled by stating a level of significance called the alpha level (usually symbolized as α). It is the largest probability of committing a Type I error that the researcher would allow and still decide to reject the null hypothesis. Type II errors occur when the null hypothesis is not rejected when, in fact, it is false.

Test instruments

1. Fisher's tests (F-test)

The *F*-test compares the spread of two sets of data to test if the sets belong to the same population; meaning if the precisions are similar or not. This test makes use of the ratio of the two variances. It is a test of significance that is used in the place of chi square test in a 2 by 2 table, especially when the sample sizes are small. The *F*-test makes use of the formula below:

$$p = ((a+b)!(c+d)!(a+c)!(b+d)!)/a!b!c!d!N!$$

In this formula, the 'a,' 'b,' 'c' and 'd' are the individual frequencies of the 2X2 contingency table, and 'N' is the total frequency. The Fisher test uses this formula to obtain the probability of the combination of the frequencies that are actually obtained. It also involves the finding of the probability of every possible combination which indicates more evidence of association.

2. T-tests

The T-test refers to a form of inferential statistics that is used to determine if there is a significant difference between the means of two groups that may be related in certain characteristics. It is a hypothesis testing tool, which enables testing of assumptions applicable to a population (Ali & Bhaskar, 2016). The T-test uses the t-distribution values and the degrees of freedom to determine the statistical significance. It allows the comparison of average values among two data sets and determine if the values are from the same population. To calculate a t-test, three data values are needed. They are; the mean difference, i.e. the difference between the mean values of each data set; the standard deviation of each group, and the number of data value of each group (Kenton, 2020).

The result of the t-test will produce the t-value, which when calculated is compared against a given value gotten from the T-Distribution table (also called the critical value table). This comparison helps measure the impact of chance on the difference alone, and whether the difference is beyond the range of chance. The t-test questions if the difference between the groups is a real difference in the sample or if it is a random difference that is potentially meaningless. There are four types of T-tests, there is the student's t-test, the Cochran's test, T-test for large data set where n is greater than or equals to 30, and the paired t-test. *The student's t-test is usually used* for comparison between two independent sets of data that have related standard deviations; the Cochran t-test is used when the standard deviation of the independent data sets reveal a significant difference, and the paired t-test is used for comparison of strongly dependent sets of data

There are four major T-test assumptions

- i. The assumption that the scale of measurement applied to the data collected for a T-test follows a continuous or ordinal scale.
- ii. The assumption that in a simple random sample, data is collected from a randomly selected representative of the total population.
- iii. The assumption that result from data plotting in a normal distribution is usually in a bell-shaped curve.

- iv. The last assumption is that there is a homogeneous variance that exists when the recorded standard deviation are approximately equal

$$T = \frac{\text{mean1} - \text{mean2}}{\frac{s(\text{diff})}{\sqrt{n}}}$$

Where:

mean1 and mean2 are the average values of each sample set.
s(diff) is the standard deviation of differences of the paired values
n is the sample size, the total number of paired differences

3. Analysis of Variance (ANOVA)

Analysis of variance is a statistical technique that is used to check if the means of two or more groups are significantly different from each other. It checks the impact of one or more factors by comparing the means of different samples. It was invented by R.A Fisher which is why it is sometimes referred to as Fisher's ANOVA.

The use of this parametric statistical technique involves certain key assumptions which include the following:

- i. Independence of case: This means that the case of the dependent variable should be independent or sample should be selected randomly. There should be no pattern in the selection of the sample.
- ii. Normality: Distribution of each group should be normal. The Kolmogorov-Smirnov or the Shapiro-Wilk test maybe used to confirm the normality of the group.
- iii. Homogeneity: It means variance between the groups should be the same. This can be tested using the Levene's method.

If a particular data set follows the above assumptions, then the analysis of variance is the best technique to compare the means of two or more populations.

There are two major types of ANOVA.

- i. One way analysis: When comparison is made between more than three groups based on one factor variable, it is said to be one way analysis of variance. For instance, if you want to compare whether or not the main output of three students is the same based on the studying hours of the three students.
- ii. Two way analysis: When factor variables are more than two, then it is said to be two way analysis of variance. For example, based on the studying hours, comparison can be made on whether or not the main output of the three students is the same.

The formula below represents one-way Anova test statistics:

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Squares (MS)	F
Within	$SSW = \sum_{j=1}^k \sum_{i=1}^l (X_{ij} - \bar{X}_j)^2$	$df_w = k - 1$	$MSW = \frac{SSW}{df_w}$	$F = \frac{MSB}{MSW}$
Between	$SSB = \sum_{j=1}^k (\bar{X}_j - \bar{X})^2$	$df_b = n - k$	$MSB = \frac{SSB}{df_b}$	
Total	$SST = \sum_{j=1}^n (\bar{X}_j - \bar{X})^2$	$df_t = n - 1$		

Table adapted from Byjus learning app.

$$F = MST/MSE$$

$$MST = SST/p-1$$

$$MSE = SSE/N-p$$

$$SSE = \sum (n-1)s^2$$

Where,

F = Anova Coefficient

MSB = Mean sum of squares between the groups

MSW = Mean sum of squares within the groups

MSE = Mean sum of squares due to error

SST = total Sum of squares

p = Total number of populations

n = The total number of samples in a population

SSW = Sum of squares within the groups

SSB = Sum of squares between the groups

SSE = Sum of squares due to error

s = Standard deviation of the samples

N = Total number of observations

1. Linear correlation and Regression

Regression analysis is a variant of statistical methods used to predict the possible relationship between a dependent variable and one or more independent variables. It is sometimes used to assess the strength of the relationship between variables and for modeling the future relationship between them. The two major and most used types of regression is the linear regression and the logistic regression. Linear Regression is a commonly used form of predictive analysis. That is, it is used to predict the continuous dependent variable using a given set of independent variables. On the

other hand logistic regression is used to predict the categorical dependent variable using a given set of independent variables. For instance linear regression outputs come for variables such as age, cost etc which are of a continuous nature, while logistic regressions are usually categorical it is either 0 or 1, yes or no etc

SUMMARY

A hypothesis is an unconfirmed statement about the likely relationship between variables. It is a clear, precise and testable prediction about what a researcher expects to find in a study. The null hypothesis stated as the H_0 is a statement about a population parameter that is assumed to be true. An alternative hypothesis is stated as H_1 is an unconfirmed statement that directly contradicts a null hypothesis by stating that the actual value of a population parameter is less than, greater than, or not equal to the value stated in the null hypothesis. The level of significance refers to the standard of judgment that has been set, upon which a decision is made regarding the value stated in a null hypothesis. A p value is the probability of obtaining a sample outcome, given that the value stated in the null hypothesis is true. The p value of a sample outcome is compared to the level of significance.

REFERENCES

- ALI, Z., & BHASKAR, S. B. (2016). Basic statistical tools in research and data analysis. *Indian journal of anaesthesia*, 60(9), 662–669. <https://doi.org/10.4103/0019-5049.190623>
- CASTILLO, M. (2013). The Scientific method: a need for something better? *AJNR Am J Neuroradiol*. 2013; 34(9):1669-71. Doi:10.3174/anjr.A3401
- CRESWELL, J.W. (1994). Research Design: Qualitative and Quantitative approaches. Thousand Oaks, CA: Sage. How to report P values in journals. - FAQ 1885 –GraphPad. <https://www.graphpad.com/support/faq/how-to-report-p-values-in-jn>
- KENTON, W (2020). Tools for Fundamental Analysis: T-Test. Reviewed by Westfall Peter. *Investopedia*. March 2020.
- LANGKOS, S. (2014). Chapter 3- Research Methodology: Data Collection method and research tools. Doi:10.13140/2.1.3023.1369.
- LAVRAKAS, P. J. (2008). *Encyclopedia of survey research methods* (Vols. 1-0). Thousand Oaks, CA: Sage Publications, Inc. doi: 10.4135/9781412963947
- OBADAN, M.I. (2012) Research Process, Report Writing and Referencing. Abuja: Goldmark Press Ltd.
- PRASAD, S., RAO, A., & REHANI, E. (2001). Developing hypothesis and research questions. *500 Research Method*, 1-30.
- ROGERS, E (1966). Chapter 1: Research Questions - Chapter 1 –LibGuides. <https://ncu.libguides.com/c.php?g=1006886&p=7294760>
- SHAYIB, M.A. (2018). Inferential Statistic. The Basics for Biostatistics: Volume II. Bookboon,com.a hypothesis is ISBN 978-87-403-2127-2
- What does it mean to "create a null hypothesis"? (n.d). Availabble at: <https://answers.yahoo.com/question/index?qid=20090325141323AAbJEU6>