STATISTICAL POWER OF HYPOTHESIS TESTING USING PARAMETRIC AND NONPARAMETRIC METHODS

BY

UMMI-KUSUM ISA AHMED
MSC/SCIE/04149/09-10

A Thesis Submitted to the Postgraduate School, Ahmadu Bello University, ZARIA, in Partial Fulfilment of the Requirement for the Award of Masters Degree in Statistics, in Department of Mathematics.

APRIL, 2015

DECLARATION

I declare that the work in this thesis titled "STATISTICAL POWER OF HYPOTHESIS TESTING USING PARAMETRIC AND NONPARAMETRIC METHODS" has been performed by me in the Department of Mathematics under the supervision of Dr. H.G. Dikko and Dr. A. Yahaya. The information derived from literature has been duly acknowledged in the text and a list of reference provided. No part of this thesis was previously presented for another degree or diploma at any University or Institution.

UMMI-KUSUM ISA AHMED

Date

CERTIFICATION

This thesis titled "STATISTICAL POWER OF HYPOTHESIS TESTING USING PARAMETRIC AND NONPARAMETRIC METHODS" by Ummi-Kusum Isa Ahmed meets the regulations governing the award of the degree of Masters of Science of Ahmadu Bello University, and is approved for its contribution to knowledge and literary presentation.

Dr. H. G. Dikko ---------------- Date------------------------------
Chairman, Supervisory Committee

Dr. A. Yahaya ---------------- Date------------------------------
Member, Supervisory Committee

Dr. B. Sanni ---------------- Date------------------------------
Head of Department

Prof. A. J. Adebayo ----------------------- Date-----------------------------------
Dean, School of Postgraduate School

DEDICATION

I dedicate this work to my parents, beloved husband and my children.

ACKNOWLEDGEMENTS

With gratitude to Almighty Allah (SWT) the giver of life and knowledge, the author and finisher of our faith, without whom there is neither a beginning nor an end, I acknowledge His utmost divine guidance throughout my M.sc programme.

I wish to acknowledge the kind attention and inspiration provided by my supervisors, Dr. H.G Dikko and Dr. A. Yahaya who both remained consistent in spite of my shortcomings, to enable me get my project done and rightly too. To all those who saw me through the class: Prof. S.U. Gulumbe and Prof. E.O. Asiribo, the Head of Department, Dr. B. Sanni, including other lecturers in the Department of Statistics, A.B.U, as well as the PG Coordinator, Dr. A. Yahaya .

I cannot measure the depth of my sincere appreciation. May Allah reward you all in greater measure. I wish to also thank my colleagues and head at the Department of Maths, Statistics and Computer Science, Kaduna Polytechnic, for their support and encouragement. My classmates: Joy, Chinene, Chollum, Safiya and others you were all very wonderful and inspiring. It has been a worthwhile encounter sharing this experience with you all.

To members of my family especially my husband, Ahmed B. Said and my children, Ammar Zaytun and Harith, thank you for your understanding when I had to be away. We shall all share in the joy of this achievement. To my parents, Alhaji and Hajiya Saidu Isa and my sisters and brother thank you for your encouragements. Also to my in laws Alhaji and Hajiya Abdullahi Ahmed thank you so much for everything may Allah reward you. And to my sister in laws and brother in law thanks so much.

ABSTRACT

Parametric and nonparametric techniques are two broad statistical methods for significance testing among continuous random variables. In this thesis, parametric and nonparametric techniques were utilized to test the power of the tests. The real-life data is simulated, generated from normal and exponential distribution. Two nonparametric tests and their parametric tests equivalents were carried out, they include; Wilcoxon Rank-Sum test and Kruskal-Wallis test as well as their parametric counterparts; independent t-Test and One-Way Anova respectively. The comparison is based on the voilation of assumption of nomality and homogenity of variance. The tests were subjected to three cases depending on the sample sizes, $n \leq 30$, and $n \geq 30$ at α= 0.05, 0.01 and0.1 significance levels. It was observed from the analysis performed at $n = 10$ and $n = 45$ for Independentent T-test and Wilcoxon Rank-Sum test under the normal distribution that the power of the test are the same that is the two tests performed equally at all levels of significants, but at n=30 the two tests perfomed equally at $\alpha = 0.05$ but at α= 0.01 and 0.1 the nonparametric is as powerful as the parametric. Under the exponential distribution, the parametric test is more powerful at $\alpha = 0.05$ and 0.1 for n = 45 and 30, but the nonparametric is more powerful for n=10, at $\alpha = 0.01$ the three size performed differently. Also under the normal distrbution for more than two independent samples, for the three sample sizes at α= 0.05 and 0.1 and also at α= 0.01 for n= 45 and 10, the Parametric test is more powerful but for n=30 the nonparametric test is as powerful as the Parametric Test. Under the exponential at the three levels for n= 45 and 30 the parametric test is more powerful but for $n = 10$ also at the three levels the nonparametric is more powerful. The power is also represented on bar chart. therefore the high chance of committing Type I orType II error is less when sample size is large and parametric test is more powerful.

TABLE OF CONTENTS

CHAPTER THREE

RESEARCH METHODOLOGY

CHAPTER FOUR

DATA PRESENTATION AND ANALYSIS

CHAPTER FIVE

DISCUSSIONS AND CONCLUSIONS

CHAPTER ONE

INTRODUCTION

1.1 Background

Nonparametric approaches are often utilized when the conditions for parametric approaches are not satisfied and in most cases when the scale of measurement is ordinal or nominal. Statistical procedure in which inferences are made about the population parameters are referred to as Parametric Statistics Cyprain (1990). Parametric approach follows certain assumptions which include samples that are randomly drawn from a normally distributed population,

1. Consist of independent observations, except for paired values,
2. Consist of values on an interval or ratio measurement scale,
3. Have respective populations of approximately equal variances,
4. Are adequately large, and
5. Approximately resemble a normal distribution.

If any of the samples breaks one of these rules, then the assumptions of a parametric test are violated. The nature of the study might be changed to adhere to the rules. If an ordinal or nominal measurement scale is being used, the study might be redesigned to use an interval or ratio scale. Also, try to seek additional participants to enlarge the sample sizes. Unfortunately, there are times when one or neither of these changes is appropriate or even possible. There are three major parametric assumption, which are and will continue to be violated by researchers in health sciences; level of measurement, sample size and normal distribution of the dependent variable Pett (1992).

If samples do not resemble a normal distribution, you might have learned to modify them so that you can use the tests you know. There are several legitimate ways to modify your data, so you can use parametric tests. First, if the reasons can be justified, then the extreme values from samples called outlier might be removed. Second, you can apply a mathematical adjustment to each value in your samples called a transformation. That is you might square every value in a sample. Transformations do not always work, however. Third, there are more complicated methods that are so advanced. Fortunately, there is a family of statistical tests that does not demand all the parameters, or listed rules above. They are called nonparametric tests.

Roughly speaking, a nonparametric procedure is a statistical procedure that has certain desirable properties that hold under relatively mild assumptions regarding the underlying population from which the data are obtained. Although nonparametric assumptions do not require the stringent assumptions associated with their parametric counter paths this does not imply that they are assumptions free Pett (1992). The rapid and continuous development of nonparametric statistical procedures over the past six decades is due to some advantages. This informs the need to compare results between parametric and nonparametric statistical procedures for some certain results.

As related to experimental designs therefore, it is imperative to combine both parametric and nonparametric approaches in the test of hypothesis, because experimental results could produce both continuous and categorical variables (ordinal and nominal variables).

Statistical inference procedures enable researchers to determine, in terms of probability, whether the observed differences between sample data could easily occur by chance or not. If sample size is very small there may be no alternative to using a nonparametric statistical test. Siegel and Castellan (1988).

However it is important to discuss the difference between parametric and non-parametric test because this leads to a major decision for the researcher– in choosing the appropriate test. In the development of modern statistics the first methods developed made a lot of assumptions about the characteristics of the population from which the samples were drawn. That is, they made assumptions about the population parameters and the test is referred to as parametric tests. The most obvious assumption is that the data were randomly drawn from a normally distributed population. Another assumption is that the data are randomly drawn from populations having the same variance. These assumptions make the general overriding assumption that the probability distribution of the population (from which the sample was drawn) is known in advance. The most common distribution assumed is the normal distribution. It has generally been argued that parametric statistics should not be applied to data with non- normal distributions empirical research has demonstrated that Wilcoxon Rank-Sum test generally has greater power than the T-test unless data are sampled from normal distribution Siegel and Castellan (1988).

More recently, distribution free or non-parametric tests have been developed and subsequently commonly used. These tests do not trigger assumptions and in particular do not have the overriding assumption of a normally distributed population.

Sometimes the nonparametric procedures are simpler than their parametric counterparts. On the contrary, a primary criticism of using parametric methods in statistical analysis is that they oversimplify the population or process we are observing. Indeed parametric tests are not more useful because they are perfectly appropriate, rather because they are perfectly convenient.

However, even when the parametric assumptions hold perfectly true, we will see that nonparametric methods are only slightly less powerful than the more presumptuous statistical methods. Furthermore, if the parametric assumptions about the data fail to hold, only the nonparametric method is valid. A t-test between the mean of two normal populations can be dangerously misleading if the underlying data are not actually normally distributed.

The other phase of statistical inference is hypothesis testing which some people feel is the more important aspect of statistical inference. Although the modern trend is to view testing statistical hypotheses from the point of decision theory, in this study we will view it from the classical point of view, having choice of accepting or rejecting a given hypothesis. Also, we will confine ourselves to two decision (action) problems. That is the `status quo' hypothesis called the null hypothesis and is denoted by Ho. The hypothesis denoting the change is called the alternative hypothesis and is denoted by H_1.

The hypothesis test comprises two mutually exclusive statement s, the alternative and the null hypotheses. The null hypothesis states the negative case, that ‘it is not true or there is no difference’, and the alternative hypothesis states that ‘it is true or there is a difference'. The procedure involved is a scientific one that is founded in simple logic for the purpose of being both open and potentially repetitive (can be replicated by others).

The following steps outline the hypothesis testing procedure:

1. State the null (H_o) and alternative (H_1) hypotheses.
2. Decide whether parametric or non parametric.
3. Choose a statistical test to test H_0.
4. Specify a significance level (alpha=α) or probability level for rejection of Ho.
5. Determine the sample size (n).
6. Assume (or find) the sampling distribution of the statistical test in 2.
7. On the basis of 2, 3, 4 and 5 above, define the region of rejection of H_o.
8. Compute the value of the statistical test using the sample data.
9. If the resultant value of the test is in the rejection area, reject H_o.
10. If the resultant value of the test is outside the rejection area H_o is not rejected at the α level.

Numerous advantages have been identified with non parametric procedures, such as: Tests are available for dealing with samples from different population; it usually depends on minimum assumptions and is thus less subject to improper use; they are the only alternative for small sample sizes unless the population distribution is known. In addition, computations are usually fast and easy to understand. However, a major disadvantage of the procedure is that, it is considered wasteful if all the assumptions for parametric test hold.

1.2 Power Efficiency

This is the amount of increase in sample size which is necessary to make a test B as powerful as test A.

The table below illustrates the relationship between type I error, type II error and power.

Table1.1. Relationship between Type I Error and Type II Error

Decision	**Null Hypothesis**	**Alternative Hypothesis**
	H_o *is* True	H_o *is* False
Reject	Type I error, α= p(Type I error)	Correct decision
Accept/Do not Reject	Correct decision	Type II error , β=p(Type II error)

Rather than referring to type II error, statisticians usually use 1- β which is called the power of a statistical test. The different probability values for β that occur when alternative is composite can be presented by a functional relationship known as power function.

Ways of increasing power.

1. Increasing α will increase power but it also increase chance of a Type 1 error.
2. Increasing sample size.
3. Using ratio or interval data versus nominal or ordinal. Test involving ratio/ interval are called parametric test. Non-parametric test are also applied on data measured on ratio/ interval scale.

4. Using repeated measures test such as the repeated measures t-test or ANOVA. By using the same subjects repeatedly, variability is reduced.
5. If variance are equal use pooled estimates of variance (e.g independent group t-test).
6. Increasing measurement precision, increase probability of finding a significant difference.
7. Using sample from extremes of the distribution. Reduces generalizability of experiment reduces variability.

Noticeable differences between parametric and non parametric procedures are,

i. In parametric approach, the conditions about population from which sample is taken are specified, while conditions for non parametric approach are fewer and weaker.
ii. Parametric procedures utilize information based on measurements, while in non parametric procedures; measurements are reduced to ranks and signs.
iii. Parametric procedure requires at least interval scale of measurement while non parametric procedure uses ordinal scale.
iv. Parametric procedures are usually based on mean, while non parametric procedures are usually based on median.

The classical parametric methods in analysis of variance, from one-way to multi-way tables, often suffer from sensitivity to the effects of non-normal data. The nonparametric methods are much more robust. In most cases, they mimic their parametric counterparts but focus on analyzing ranks instead of response measurements in the experimental outcome.

In practice, for a random variable X, parametric approach can be used if the following conditions are satisfied:

1. The underlying distribution of the population is known.
2. The underlying distribution satisfies the normality assumptions. That is;
   - Error term is a random vector.
   - Error term has mean zero.
   - Constancy of error variance (Homoscedasticity).
   - Error terms are uncorrelated. (No serial or autocorrelation)
   - The error terms are Normally Independent Identical Distribution (NIID) $(0, \sigma^2)$.
3. The scale of measurement is at least interval.

In cases when these assumptions are not satisfied, the parametric method can be criticized regarding validity and optimality. Then the non parametric approach will be used. Non parametric methods based on ranks are valid for some broad family of underlying distributions. It is however often argued that except for a simple design such as matched pair or completely randomized design their powers tend to be low and their possibilities to detect different hypotheses are limited. For example in a randomized complete block design Friedman's test can be used to test for differences between the treatments. Since it is based on intra- block ranking, its sensitivity is low particularly if the number of observation for each block is small.

Non parametric tests are distribution free tests that do not depend on the distribution of the population from which the sample used is taken. It can be used when:

1. The conditions for parametric test are not satisfied.
2. Ordinal scale of measurement is used.

TABLE1.2: Comparison of Statistical Tests

DATA TYPE	**PARAMETRIC RATIO, INTERVAL**	**NONPARAMETRIC ORDINAL**	**NORMINAL**
Single sample	Z-test ,T-test	Sign test, K-S test	X^2 Goodness of fit
Two independent sample	Z-test, T-test	Wilconxon Rank sum (Mann Whitney u)	
Two dependent sample	Paired T-test	Paired sign Wilcoxon signed-rank	
Two factors	Two- way ANOVA	Fredman test	X^2 Test of independent
Comparison of several means	One- way ANOVA	Kruskal –Wallis test	

1.3 Motivation

In modern statistical methods a good number of assumptions are made about the nature of population from which samples are drawn and data were collected. These statistical techniques are also known as parametric tests.

On the other hand, development of a large number of techniques of inferences which do not make stringent assumptions about the population from which samples are drawn is of great interest. These techniques are known as nonparametric statistical techniques.

The nonparametric methods make less of many assumptions which are made under parametric statistical tests; which makes them easy to apply. On the other hand, parametric methods are said to be more efficient in most cases. The motivation is to compare the results of these two techniques using the power of a test to examine the simplicity and efficiency between the techniques

1.4 Aim and Objectives of the Study

The study/research is aimed at investigating the Statistical Power of Hypothesis Testing using Parametric and Nonparametric Method with a view to achieving the following objectives to compare;

i. The decision of parametric test with non parametric test when normal or exponential distribution is used for simulation.

ii. The power of parametric with non parametric approaches in test of hypotheses using small and large sample size.

iii. The consistency of the two approaches in hypothesis testing.

1.5 Significance of the Study

This research work is based on hypothesis test where comparisons of the class of test were subject to two cases which are the sample size $n \leq 30$ and $n \geq 30$ in order to know the relative consistency and power of the statistical techniques. At the end of the study the research will be a goldmine to statisticians, engineers, agricultural and researchers. It has relevance for statistical approaches which consider comparing different unique cases.

1.6 Scope of the Study

This thesis is focused on comparing parametric with non parametric approaches by analyzing two sample problems and completely randomized design (CRD) using Wilcoxon-Rank Sum and Kruskal-Wallis respectively. The parametric counterpart will be student’s T- test and One-Way Analysis of Variance.

1.7 Source of Data

The data used for this research work were simulated data which were generated from the real life data which follows a normal distribution $N(0,1)$ and exponential distribution The various tests were done using SPSS 20.

CHAPTER TWO

LITERATURE REVIEW

2.1 Introduction

This chapter present review of literature on parametric and nonparametric tests by different scholars in attempt to present the most powerful among the two classes of statistical test methodologies.

2.2 The Power of a Test

The power of a statistical test is the probability that it will correctly lead to the rejection of a false null hypothesis Greene, (2000). The statistical power is the ability of a test to detect an effect, if the effect actually exists Green, (2000). Cohen (1988) says, it is the probability that it will result in the conclusion that the phenomenon exists.

Siegel (1956), Runyon (1977) and Goon and Gupta (2003) stated that a particular test (A) may be more powerful than its counterpart test (B) when their sample sizes are equal but its counterpart (B) may be as powerful as its efficiency (A) if its sample size is increased.

Goon and Gupta (2003) stated that the power of a statistical hypothesis test depends on the following factors:

- The power depends on the population standard deviation: The smaller the population standard deviation, the greater the power.
- The power depends on the sample size used: The larger the sample, the greater the power.

- The power depends on the level of significance of the test: The smaller the level of significance, the smaller the power.

Goon and Gupta (2003), consistency is one such criterion which is particularly useful in non-parametric methods. A test which is consistent for a subclass of alternatives is said to be specially sensitive to that type of alternatives and can be recommended for use in detecting differences of the type mentioned in the alternatives. They defined consistency as "The sequence of tests corresponding to $\{W_n\}$ is consistent if for every value of θ lying in $(\Theta - \Theta_0)$, the power, $P_\theta\ (W_n)$, tends to 1 as n tends to ∞. The idea behind the concept of consistency can be viewed from considering the behaviour of a test for increasing *n*. Naturally one would like that the power should increase with the number of random variables taken into account. In cases X_1, X_2, ..., X_n form a random sample, for instance, our test should have, for large sample sizes, a high probability of rejecting a false hypothesis that is a high power.

Goon and Gupta (2003) stated that another useful criterion is the concept of asymptotic relative efficiency (ARE). In point estimation, the efficiency of two unbiased estimators for a parameter is defined as the inverse ratio of their variances. In the case of power efficiency, Pitman (1948) defined it as:

> Let T_n and T_n*, n = 1, 2, ..., be two sequences of test statistics of the same null hypothesis H_0 at the same significance level α. Let the distributions be indexed by a real parameter γ, so that $\gamma = 0$ gives a distribution in H_0 and other γ's correspond to distributions native hypothesis. Considering a sequence of alternatives $\gamma_{i,}$ If, for the same power with respect to the same alternative γ_i, the test T_n requires n_i observations and the test T_n* requires

n_i* observations, then the relative efficiency of test T_n with respect to test T_n* is given by the ratio $e = n_i^*/n_i$.

Goon and Gupta (2003) stated that in general, e depends on α, γ_i and n_i*. The evaluation of $e = e(\alpha, \gamma_i, n_i^*)$ as a function of the three arguments is not simple. Some asymptotic value of $e(\alpha, \gamma_i, n_i^*)$ may be computed keeping one arrangement constant and letting the others approach some suitable limits. If the tests T_n and T_n* are consistent, then their powers will approach 1 with increasing sample size. Then let the sequence γ_i be such that $\gamma_i \to 0$ as $i \to \infty$, and such that the power of each test lies in the open interval (α, 1) for finite sample sizes and approaches some limit between α and 1. Hence, *ARE* can be defined as:

> Let T_n and T_n* be two sequences of level-α tests of a null hypothesis $\gamma = 0$ against an alternative hypothesis γ_i. The asymptotic relative efficiency (ARE) of the test T_n relative to the test T_n* is the limiting value of e while simultaneously $n_i^* \to \infty$ and $\gamma_i \to 0$, that is *ARE* = $\lim_{\substack{n_i^* \to \infty \\ \gamma_i \to 0}} \frac{n_i^*}{n_i} = \lim_{\substack{n_i^* \to \infty \\ \gamma_i \to 0}} e(\alpha, \gamma_i, n_i^*)$, if this limit exists and is constant for all increasing sequences of sample size n_i, n_i*.

They also stated that in many important applications ARE does not depend on α and ARE is also called the local asymptotic efficiency (LAE) since it is the large-sample power in the vicinity of the null hypothesis H_0. Further they stated that the criteria of consistency and ARE are both large-sample criteria and hence may not be of much use in finite-sample cases. But owing to the difficulty of exact power calculation with non-parametric tests, the above two criteria have found use in non-parametric inference. Also, in some cases ARE provides an approximation to the exact efficiency. There are some theorems with proof for the evaluation of ARE.

Breheny (2012) stated that suppose test 1 requires n_1 observations to obtain a certain power β, and that test 2 required n_2 observations to reach the same power then the relative efficiency to test 2 is n_2/n_1. He further gave an example that, if test 2 needs 100 observations to achieve 80% power and test 1 needs only 50, then the relative efficiency is n_2/n_1 = 100/50 =2; in other words, test 1 is twice as efficient. He stated that in all but the simplest situations, this is impossible to calculate instead, we must rely on either simulations or asymptotic arguments. Also, he observed that asymptotic arguments are not entirely straightforward, either, because the power of almost any test goes to 1 as $n \to \infty$. Hence, to avoid the power of both tests going to 1, we must have $\theta \to \theta_0$ as $n \to \infty$.

2.3 The Concept of Parametric and Nonparametric Tests

Most of the standard methods of statistical inference are based on the familiar assumption that the random variables have normal distributions. Then the given procedures are optimum. But for non-normal distributions the standard procedures may be far from optimum. In such cases nonparametric methods are used. The nonparametric methods are concerned with the treatment of standard problems when the assumption of normality is replaced by quite general assumptions about the populations. Classical methods based on specific population assumptions may be termed parametric methods. A procedure will be called distribution-free if the statistic used has a distribution which does not depend on the distribution function of the population from which the sample is drawn. So in such procedures assumptions regarding the population are not necessary Goon and Gupta (2003).

They stated the term nonparametric is due to Wolfowitz who used it to indicate that the population could not be specified by a finite number of parameters. A nonparametric test is concerned with the form of the population and not with any parametric values. Thus ‘distribution free’ and ‘nonparametric’ are not actually synonymous terms. They stated that the former is used to indicate the nature of the distribution of the (test) statistic and the later the type of (hypothesis) problem investigated. Also, they further stated that the field of nonparametric statistics is extensive. It is not exclusively concerned with tests of hypotheses. It may be used to estimate suitably chosen population characteristics (such as the quartiles). From theoretical considerations, it is convenient to assume that the random variables have continuous distribution functions. Theoretically, then, it is unnecessary to deal with ties. Yet ties do occur and there are practical rules for handling them.

Simon (2008) stated that a parametric test, of course, is a test that requires a parametric assumption, such as normality. A nonparametric test does not rely on parametric assumptions like normality. Further he stated that whether to choose a parametric test is a matter of judgment. But one should keep in mind several things like;

1. A nonparametric test protects against some violations of assumptions and not others. The two sample t-test requires three assumptions, normality, equal variance and independence. The nonparametric alternative, the Mann-Whitney-

Wilcoxon test, does not rely on the normality assumption, but one should better make sure that the equal variance and independence assumption is met.

2. One can often use a transformation to better match some of the assumptions like normality and equal variances. The log transformation can sometimes give you data that is much better behaved.

3. One might select a statistic on the basis of what others in the same field have used. But the question becomes, is there a lot of precedent for the type of data collected to be non-normal?

4. Is one comfortable with the fact that the non-parametric test may be evaluating a different measure than the parametric test? The Mann-Whitney-Wilcoxon test, for example, provides an estimate of $p[X > Y]$, probability that a randomly selected patient from your first population has a larger value than a randomly selected patient from the second population. This may be more interesting (or it may be less interesting) than a two-sample t-test which provides you with an estimate of the difference in the average between the first and the second populations.

Vickers (2005) stated that the rationale for recommending non-parametric over parametric methods, unless certain conditions are met, is rarely made explicit. But techniques for statistical inference from randomized trials can only fail in one of two ways: they can inappropriately reject the null hypothesis of no difference between groups (false positive or Type I error) or inappropriately fail to reject the null (false negative or Type II error). Hence any recommendation to favour one technique over another must be based on the relative rates of these two errors. Thus concern over the relative advantages of parametric and non-parametric methods has focussed on Type II error Sawilowsky and Blair (1992). Zimmerman and Zumbo (1990), Bridge and Sawilowsky (1999) noted that typically, researchers have created a large number of data sets, in which observations were created from a distribution incorporating a difference between groups. Each data set is then analyzed by both parametric and non-parametric methods in order to calculate the proportion of times the null hypothesis is rejected (that is, the power).

Vickers (2005) stated that there are two implications for methodology research on the relative value of parametric and non-parametric techniques. First, we should worry about the distribution of change scores. It seems likely that change from baseline would approximate more closely to a normal distribution than the post-treatment score. This is because change scores are a linear combination and the Central Limit Theorem therefore applies. As a simple example, imagine that baseline and post-treatment score were represented by a single throw of a die. The post-treatment score has a flat (uniform) distribution, with each possible value having an equal probability. The change score has a more normal distribution: there is a peak in the middle at zero – the chance of a zero change score is the same as the chance of throwing the same number twice, that is 1 in 6 – with more rare events at the extremes – there is only a 1 in 18 chance of increasing or decreasing score by 5.

Bradley (1996) stated that the terms nonparametric and distribution-free are not synonymous. A nonparametric test is a test that makes no hypothesis about the value of a parameter in a statistical density function, whereas a distribution-free test is one which makes no assumptions about the precise form of the sampled population. But non parametric tests are distribution free tests that do not depend on the distribution of the population from which the sampled used are taken. Related samples involve subjects providing scores for both samples, whereas with independent samples each subject contributes a score to only one sample. The two sample t-test are calculated depending on whether the two samples are related or independent Hinton (2004). Fisher and Van Belle (1993) stated that a family of probability distribution is nonparametric if the distributions of the family cannot be conveniently characterized by a few parameters. Statistical procedures that hold or are valid for a nonparametric family of distributions are called nonparametric statistical procedures.

Hinton (2004) the basic assumptions of the t-test are the same whichever t-test to be applied. The sampling distribution is to be normally distributed so, the assumption becomes; the samples came from normally distributed populations. Fortunately, the t-test is robust so that even if the distributions are only vaguely normal, humped in the middle and tailing off to the sides, the t-test is still likely to be valid. This is especially true for large samples (> 30). Further, Hinton (2004) stated that it must be assumed that the samples are randomly chosen from their populations so that we can use sample statistics (mean and standard deviation) as unbiased estimates of the population parameters. Many testing procedures based on ranks have been proposed since 1930's with perhaps the Wilcoxon test as the most well known method. These tests are appealing because they are distribution free under the null hypothesis, simple to compute, often have good efficiency compared to the classical testing procedure. Individual rank tests were proposed much earlier; the modern development of the subject begun with papers by Hotelling and Pabst

(1936), Smirnov (1936), Friedman (1937) and Kendal (1938). It would not be too much of exaggeration to say that for every parametric test there is a nonparametric analogue that allows some of the assumptions of the parametric test to be relaxed Siegel, (1956), Hollander and Wolfe, (1973).

In terms of selecting a statistical test, Campbell (2006) stated that the most important question is "what is the main study hypothesis?" In some cases there is no hypothesis; the investigator just wants to "see what is there". For example, in a prevalence study there is no hypothesis to test, and the size of the study is determined by how accurately the investigator wants to determine the prevalence. If there is no hypothesis, then there is no statistical test. It is important to decide a priori which hypotheses are confirmatory (that is, are testing some presupposed relationship), and which are exploratory (are suggested by the data). No single study can support a whole series of hypotheses. Further, Campbell (2006) stated that a sensible plan is to limit severely the number of confirmatory hypotheses. Although it is valid to use statistical tests on hypotheses suggested by the data, the P values should be used only as guidelines, and the results treated as very tentative until confirmed by subsequent studies. A useful guide is to use a Bonferroni correction, which states simply that if one is testing n independent hypotheses, one should use a significance level of 0.05/n. Thus if there were two independent hypotheses a result would be declared significant only if $P<0.025$. Note that, since tests are rarely independent, this is a very conservative procedure - one unlikely to reject the null hypothesis. The investigator should then ask "are the data independent?" This can be difficult to decide but as a rule of thumb results on the same individual, or from matched individuals, are not independent. Thus results from a crossover trial, or from a case control study in which the controls were matched to the cases by age, sex and social class, are not independent, the analysis should reflect the design, and so a matched design should be followed by a matched analysis. Also, results measured over time require special care. One of the most common mistakes in statistical analysis is to treat correlated variables as if they were independent. For example, suppose we were looking at treatment of leg ulcers, in which some people had an ulcer on each leg. We might have 20 subjects with 30 ulcers but the number of independent pieces of information is 20 because the state of ulcers on each leg for one person may be influenced by the state of health of the person and an analysis that considered ulcers as independent observations would be incorrect. It would seem prudent to use non-parametric tests in all cases, which would save one the challenge of testing for Normality. According to Campbell, (2006) parametric tests are preferred, however, for the following reasons:

1. There is no interest in a significance test alone; rather they are made about the population from which the samples came, and this is best done with estimates of parameters and confidence intervals.
2. It is difficult to do flexible modeling with non-parametric tests, for example allowing for confounding factors using multiple regressions.

2.4 Empirical Comparisons of Parametric and Non-parametric Tests

Peter and Brien (1984) considered five procedures for the comparison of two or more multivariate samples where the nonparametric rank-sum test and the generalized least square test were used and they observed that in all simulations the nonparametric procedure provided relatively good power and accurate control over the size of the test, and is recommended for general use. Alternatively, the generalized least squares procedures may also be useful with normally distributed data in moderate or large samples. Typically, researchers have created a large number of data sets, in which observations were created from a distribution incorporating a difference between groups. Each data set is then analyzed by both parametric and nonparametric methods in order to calculate the proportion of times the null hypothesis is rejected.

Bridge and Sawilowsky (1999) concluded that the t-test was more powerful only under a distribution that was relatively symmetric, although the magnitude of the difference was trivial. In contrast, the Mann-Whitney held huge power advantages for data sets which presented skewness.

Andrew *et al.* (2011) argued that analysis of covariance is an effective method for addressing two considerations for randomized clinical trials. One is reduction of variance of estimates of treatment effects and thereby the production of narrower confidence intervals and more powerful statistical tests. The other is the clarification of the magnitude of treatment effects through adjustment of corresponding estimates for any random imbalances between the treatment groups with respect to the covariables. The statistical basis of covariance analysis can be either nonparametric, with reliance only on the randomization in the study design or parametric through a statistical model for a postulated sampling process. For non-parametric methods, there are no formal assumptions for how a response variable is related to the covariables, but strong correlation between response and covariables is necessary for variance reduction. Computations for these methods are straight forward through the application of weighted least squares to fit linear models to the differences between treatment groups for the means of the response variables and the co-variables jointly with a specification that has null values for the differences that correspond to the covariables moreover, such analysis is similarly applicable to dichotomous indicators, ranks or integers for ordered categories, and continuous measurements. Since non parametric covariance analysis can have many forms, the ones which are planned for a clinical trial need careful specification in its protocol.

Further, Andrew *et al.* (2011) stated that a limitation of nonparametric analysis is that it does not directly address the magnitude of treatment effects within subgroups based on the covariables or the homogeneity of such effects. For this purpose, a statistical model is needed. When the response criterion is dichotomous or has ordered categories, such a model may have a non-linear nature which determines how covariance adjustment modifies results for treatment effects. Insight concerning such modification can be gained through their evaluation relative to nonparametric counterparts. Such evaluation usually indicates that alternative ways to compare treatments for a response criterion with adjustment for a set of covariables mutually support the same conclusion about the strength of treatment effect.

Vickers (2005) stated that it has generally been argued that parametric statistics should be applied to data with non-normal distributions. Such trials should be analyzed using ANCOVA, rather than t-test. Vickers (2005) carried out simulation studies to compare the power of Mann-Whitney and ANCOVA, for analyzing each distribution, varying sample size, correlation and type of treatment effect (ratio or shift). The results showed that change between skewed baseline and post-treatment data tended towards a normal distribution. ANCOVA was generally superior to Mann-Whitney in most situations, especially where log-transformed data were entered into the model. The estimate of the treatment effect from ANCOVA was not importantly biased. In his conclusions, he stated that ANCOVA is the preferred method for analyzing randomized trials with baseline and post-treatment measures.

Parametric methods, such as the t-test for the analysis of randomized trials can only be used for the data that are approximated to a normal distribution. Atman (1991) stated that parametric methods require the observations within each group to have an approximately normal distribution. If the raw data do not satisfy these conditions, a nonparametric method should be used, in some cases, central limit theorem is invoked such that parametric methods are said to be applicable if sample size is suitably large for reasonably large sample (say 30 or more observations in each sample). The t-test may be computed on almost any set of continuous data. Atman (1948) for example, state that "parametric methods require the observations within each group to have an approximately normal distribution, if the raw data do not satisfy these conditions a non-parametric method should be used".

a. In some cases, central limit theorem is invoked such that parametric methods are said to be applicable if sample size is suitably large: "for reasonably large samples (say, 30 or more observations in each sample) the t-test may be computed on almost any set of continuous data".

b. The rationale for recommending non-parametric over parametric methods, unless certain conditions are met, is rarely made explicit. But techniques for statistical influence from randomized trials can only fail in one of two ways; they can inappropriately reject the null hypothesis of no difference between groups false positive or (Type I error) or inappropriately fail to reject the null hypothesis (Type II error). Hence any recommendation to favor one technique over another must be based on the relative rates of these two errors. Empirical statistics research has clearly demonstrated that the t-test does not inflate Type I error, Andrew, *et al* (2011).

c. They found that where there was truly no difference between groups, Siegel and Castellan (1988), the t-test would reject the null hypothesis close to 5% of the time. Thus, concern over the relative advantages of parametric and nonparametric methods has focused on Type II error.

Techniques applicable for the analysis of longitudinal data when the response variable is non-normal are not nearly as comprehensive as for normally distributed outcomes. However, there have been several recent developments. Semi-parametric and nonparametric methodologies for the analysis or repeated measurements were reviewed Davis, (1991). The commonly encountered design, in which for each subject, one assesses a variable response at multiple fixed time points, is considered. The types of outcomes considered include binary, ordered categorical, and continuous (but extremely non-normal) response variables.

Whitehead and Whitehead (1996) presents a general parametric approach, which utilizes efficient score statistics and Fisher's information, and relates this to different methods suggested by other authors. Normally distributed, binary, ordinal and survival data were considered. Both the fixed effects and random effects model for treatments are described.

Briggs *et al.* (1997) described clinical trials which are to be analyzed with the Mantel Haenzel (logrank) test for comparing survival curves. The results allow for loss to follow up during the trial and let the planner reduce the total number of patients required by introducing a period of continued observation after the end of patient accrual. For simplicity, the actual calculation of the required trial length was based on use of the following parametric test; assuming time to death is exponential, the ratio of the maximum likelihood estimators of the hazard rates is equal to one: a method was developed to compute the approximate trial length required to assure a desired statistical power for a given significance level, hazard ratio, accrual rate, loss of follow-up rate, and length of the period of continued observation. Tables were included which gives the required trial length for a wide range of cases. A figure was included which depicts the effects on the required trial length of loss to follow-up and demonstrates the practical advantage of a period of continued observation.

Steinijans and Dilette (1998) found that 95% confidence limits for expected bioavailability can be obtained by classical analysis of variance (ANOVA). If symmetry of the confidence interval is desired about zero (differences) or unity (ratios) rather than about the corresponding point estimator, Westlake's modification can be used. Two nonparametric methods and their adoptions to bioavailability ratios were reviewed, one based on Wilcoxon's signed rank test (Tukey), and the other on Pitman's permutation test. The necessary assumptions and the merits of these procedures are discussed. The methods were illustrated by an example of a comparative bioavailability study.

Abebe (2009) tested the efficiency of the nonparametric one-way Kruskal-Wallis test to the parametric one-way analysis of variance (ANOVA) F-test using an experimental design data presented in Montgomery (2005). The design is on the tensile strength of a synthetic fibber used to make cloth for men's shirts which is of interest to a manufacturer. It was suspected that the strength is affected by the percentage of cotton in the fibber. Five levels of cotton percentage were considered: 15%, 20%, 25%, 30% and 35%. For each percentage of cotton in the fibber, strength measurements (time to break when subject to a stress) are made on five pieces of fibre. The stated null hypothesis (H_0) was that the mean strength of each of the five pieces of fibber are the same as against the null hypothesis which state that at least one of the means are different. It was found that both the one-way Kruskal-Walis test and the one-way ANOVA F-test resulted in the same conclusion in which the null hypothesis was rejected at $\alpha = 0.01$. Abebe (2009) stated that the one-way Kruskal-Walis test is a powerful equivalent once non-normality is suspected.

CHAPTER THREE

RESEARCH METHODOLOGY

3.1 Introduction

In this section, the different test statistics for the parametric and non parametric approaches were explained in detail. The data used for the research is from the test on an improved designed threshing machine for sorghum obtainable from Division of Agric college (DAC) Zaria. The speed rate for the machine blowing fan is at three levels after threshing the sorghum the percentage of mechanical grain damage (MGD) was estimated. Two parametric tests and their nonparametric tests equivalents; the independent t-test and one-way ANOVA their nonparametric equivalents are Wilcoxon Rank-Sum and Kruskal-Wallis test respectively.

The data is simulated fifty times using normal and exponential distribution. The equality sample number of the case $n = 10$, n=30 and $n = 45$ were used for the study. The comparison is done using different significant level $\propto = 0.05$, 0.01 and 0.1. The simulated data is used for the comparison, the simulated data was generated using SPSS 20 and the various tests were done using SPSS 20. Also a bar chart is used to represent the power of the test.

3.2 Classical Comparison between Parametric and Nonparametric Tests

In order to examine the possible effects of violation of assumptions using parametric procedures, this study is an exploratory investigation into the use of parametric versus nonparametric procedures using a single case study design. The case study investigation guidelines outlined by Conover (1999) served as the methodology.

The following univariate statistical procedures, identified as the most commonly employed procedures in the occupational therapy literature (from 1980 to the present-time), served as the basis of the investigation: (1) independent t-test; (2) paired t-test; (3) one factor analysis of variance; (4) Pearson product moment correlation; and (5) regression. In each case investigated, meaningful patterns or characteristics of the data emerged which could aid interpretation regardless of class of data analysis procedure (parametric versus nonparametric) employed. Further, information on distributional characteristics of variables as revealed by exploratory data analysis compiled from research investigation over time served to define characteristics of variables studied in any research.

3.3 Testing for Equality of Two Independent Sample

Hypothesis

$H_0 : t_1 = t_2 \; Vs \; H_1 : t_1 \neq t_2$

Where

t_1 = the population mean

t_2 = the hypothesized mean

3.3.1 Wilcoxon Rank-Sum Test

This is used for comparing a new treatment or procedure with a known standard method. For the one-sample Wilcoxon test, the procedure is to subtract the theoretical μ_0 and rank the differences according to their numerical value ignoring the sign, and then calculate the sum of the positive or negative ranks. The point is that assuming only that the distribution is symmetric around μ_0, the test statistic corresponds to selecting each number from 1 to *n* with probability $\frac{1}{2}$ and calculating the sum. The distribution of the test statistic can be calculated exactly, at least in principle. It becomes computationally excessive in large samples, but the distribution is then very well approximated by a normal distribution. There is not quite as much output as from t.test, due to the fact that there is no such thing as a parameter estimate in a nonparametric test and therefore no confidence limits, etc. either. Actually, it is possible under some assumptions to define a location measure and calculate confidence intervals for it.

The relative merits of distribution-free (or nonparametric) versus parametric methods like the *t* test are a contentious issue. If the model assumptions of the parametric test are fulfilled, then it will be somewhat more efficient, on the order of 5% in large samples, although the difference can be larger in small samples. Notice, for instance, that unless the sample size is 6 or above, the signed-rank test simply cannot become significant at the 5% level. This is probably not too important, though; what is more important is that the apparent lack of assumptions for these tests sometimes misleads people into using them for data where the observations are not independent or where a comparison is biased by an important covariate. The Wilcoxon tests are susceptible to the problem of *ties*, where several observations share the same value. In such cases you simply use the average of the tied ranks; for example, if there are four identical values corresponding to places 6 to 9, they will all be assigned the value 7.5. This is not a problem for the large-sample normal approximations, but the exact small-sample distributions become much more difficult to calculate and Wilcoxon test cannot do so.

Assumptions

1. The two samples are independent.
2. The variables observed are continuous.
3. The scale of measurement is at least ordinal.
4. The two populations differ only in location.

Let N be the sample size, the number of pairs. Thus, there are a total of *2N* data points.

For $i = 1.2.3 \dots . N$, let x_{1i} *and* x_{2i} denote the measurements.

H_0: median difference between the pairs is zero

H_1: median difference is not zero.

1. For $i = 1.2.3 \dots . N$, calculate $|x_{2i} - x_{1i}|$ and $sgn\ (x_{2i} - x_{1i})$, where sgn is the sign function.
2. Exclude pairs with $|x_{2i} - x_{1i}| = 0$ Let N_r be the reduced sample size.
3. Order the remaining N_r pairs from smallest absolute difference to largest absolute difference, $|x_{2i} - x_{1i}|$
4. Rank the pairs, starting with the smallest as 1. Ties receive a rank equal to the average of the ranks they span. Let R_i denote the rank.
5. Calculate the test statistic W,

$$W = |\sum_{i=1}^{N_r}[sgn(x_{2i} - x_{1i}). R_i]|$$

The absolute value of the sum of the signed ranks

6. As N_r increases, the sampling distribution of W converges to a normal distribution. Thus,

For $N_r \geq 10$, a z-score can be calculated as

$$Z = \frac{W - 0.5}{\sigma_W} , \sigma_W = \sqrt{\frac{N_r (N_r + 1)(2N_r + 1)}{6}}$$

If $Z > Z_{critical}$ then reject H_0

For $N_r < 10$, W is compared to a critical value from a reference table

$If\ W > W_{critical}\ then\ reect\ H_0$

Alternatively, a *p*-value can be calculated from enumeration of all possible combinations of W given N_r.

3.3.2 Parametric Approach

The classical test for comparing two treatment means is student's t-test. The two-sample *t* test is used to test the hypothesis that two samples may be assumed to come from distributions with the same mean. The theory for the two- groups sample *t* test is not very different in principle from that of the one-sample test. Data are now from two $x_{11,\ldots\ldots} x_1\, n_1 and\ x_{21,\ldots\ldots} x_2\, n_2$, which we assume are sampled from the normal distributions $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$ and it is desired to test the null hypothesis $\mu_1 = \mu_2$.

then

$$t = \frac{\bar{X}_1 - \bar{X}_2}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \quad ; \quad t_{\frac{\alpha}{2},\ n_1+n_2-2},$$

is calculated. Where

$$\sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}} = S_p$$

An alternative procedure due to Welch is to calculate the SEMs (standard error of means) from the separate group standard deviations S_1 and S_2, with this procedure, *t* is actually not *t*-distributed, but its distribution may be approximated by a *t* distribution with a number of degrees of freedom that can be calculated from S_1 , S_2 , and the group sizes. This is generally not an integer. The Welch procedure is generally considered the safer one. Usually, the two procedures give very similar results unless both the group sizes and standard deviation are very difference.

The test statistic t follows student t – distribution with $n_1 + n_2 - 2$ degrees of freedom under the null hypothesis. Under the alternative (that is when H_0 is rejected), the test statistic t follows non central t – distribution with parameter α, where α, is the non centrality parameter (NCP). That is $t \sim t\alpha_{/2,n_1+n_2-2}$ under a true null hypothesis (H_0) and $t \sim t\alpha_{/2,n_1+n_2-2}$, α under a false null hypothesis (H_1).

Where $t\alpha_{/2,n_1+n_2-2}$ is determined from *t*-distribution with $n_1 + n_2 - 2$ degrees of freedom.The calculated statistic *t* follows student's *t* distribution when the null hypothesis is true but it follows a non-central *t* distribution when the null hypothesis is false with non centrality parameter α.

Since student's t-test and Wilcoxon rank-sum test were developed for different situation, the t-test under the assumption of normality and Wilcoxon test without this assumption, there appears to be point in comparing them.

However, situations commonly arise in which it is not clear about the tests that are more appropriate. To see which of the two tests is more appropriate in such situations, it is necessary to compare the two approaches by making use of their power function.

3.4 Testing For Equality of Means In Single Factor Experiment

Model; $y_{ij} = \mu + t_i + e_{ij}$, $i = 1,2,\ldots,t$

Hypothesis

$H_0 : t_i = 0$ Vs $H_1 : t_i \neq 0$

Where

μ is the overall mean

t_i is the i^{th} treatment effect and

e_{ij} is the random error component. e_{ij}~NIID $(0,\sigma^2)$

3.4.1 Kruskal-Wallis Test

Kruskal-Wallis test is used for comparing several means whether there is any difference among the treatments, If treatments differ widely among each other, one would expect big differences among the level of response of treatments. The Kruskal-Wallis statistic is given as

$$H = \frac{12}{n(n+1)} \sum_{i=1}^{k} \frac{R_j^2}{n_i} - 3(n+1)$$

R_j is the sum of ranks assigned to observations in the ith group, for i= 1,2......k. n_i is the number of observation in the ith group; and n = $\sum_{i=1}^{k} n_i$; is the number of observation in all the k groups combined.

Under the null hypothesis that the sampled populations are identical and if each sample consists of at least five observation , then the test statistics has approximately a chi-square distribution with (k-1) degrees of freedom. Hence the null hypothesis is rejected at the α level of significance if $H_1;X^2_{1-\propto\,;k-1,}$ and is accepted, otherwise, where $X^2_{1-\propto\,;k-1}$ is the critical chi-square value.

The Kruskal-Wallis (KW) test is a logical extension of the Wilcoxon-Mann-Whitney test. It is a nonparametric test used to compare three or more samples. It is used to test the null hypothesis that all populations have identical distribution functions against the alternative hypothesis that at least two of the samples differ only with respect to location (median), if at all.

The Kruskal-Wallis test is the analogue to the F-test used in the one-way ANOVA. While analysis of variance tests depend on the assumption that all populations under comparison are independent and normally distributed, the Kruskal-Wallis test places no such restriction on the comparison. Suppose the data consist of *k* independent random samples with sample sizes *n.* Let n =$n_1 + n_{2\ldots\ldots\ldots} + n_k$ under the null hypothesis, we can claim that all of the *k* samples are from a common population. The expected sum of ranks for the sample i**,** E (R,), would be *n,* times the expected rank for a single observation. That is, *n,(n* +1)/2, and the variance can be calculated as Var(R,) = n(n + 1)(n +2)/12.

ASSUMPTIONS

1. Populations are continuous and identical.
2. Treatments are independent.

Methods

1. Rank all data from all groups together; i.e., rank the data from 1 to N ignoring group membership. Assign any tied values the average of the ranks they would have received had they not been tied.
2. The test statistic is given by:

$$K = (N-1)\frac{\sum_{i=1}^{g} n_i\ (\bar{r}_{i.} - \bar{r})^2}{\sum_{i=1}^{g}\sum_{j=1}^{n_i}\left(r_{ij} - \bar{r}\right)^2}$$

Where: r_i is the number of observations in group i

r_{ij} is the rank (among all observations) of observation j from group i

N is the total number of observations across all groups

$$\bar{r}_{i.} = \frac{\sum_{j=1}^{n_i} r_{ij}}{n_i}$$

$\bar{r} = \frac{1}{2}(N+1)$ is the average of all the r_{ij}.

3. If the data contain no ties the denominator of the expression for K is exactly $(N-1)N(N+1)/_{12}$

And

4. $\bar{r} = \frac{N+1}{2}$. Thus

$$K = \frac{12}{N(N+1)} \sum_{i=1}^{g} n_i \left(\bar{r}_{i.} - \frac{N+1}{2}\right)^2$$

$$= \frac{12}{N(N+1)} \sum_{i=1}^{g} n_i \; \bar{r}_{i.}^2 - 3(N+1)$$

The last formula only contains the squares of the average ranks.

5. A correction for ties if using the short-cut formula described in the previous point can be made by dividing K by

$$1 - \frac{\sum_{i=1}^{G} (t_i^{\,3} - t_i)}{N^3 - N}$$

Where G is the number of groupings of different tied ranks, and t_i is the number of tied values within group i that are tied at a particular value. This correction usually makes little difference in the value of K unless there are a large number of ties.

6. Finally, the p-value is approximated be $P_r(\chi^2_{g-1} \geq K)$.

If some $n_i$values are small (i.e., less than 5) the probability distribution of KS can be quite different from this chi-squared distribution. If a table of the chi-squared probability distribution is available, the critical value of chi-squared, $\chi^2_{\alpha:\,g-1}$, can be found by entering the table at $g-1$ degrees of freedom and looking under the desired significance or alpha level. The null hypothesis of equal population medians would then be rejected if $K \geq \chi^2_{\alpha:\,g-1}$. Appropriate multiple comparisons would then be performed on the group medians.

7. If the statistic is not significant, then there is no evidence of differences between the samples. However, if the test is significant then a difference exists between at least two of the samples. Therefore, a researcher might use sample contrasts between individual sample pairs, or post hoc tests, to determine which of the sample pairs are significantly different. When performing multiple sample contrasts, the Type I error rate tends to become inflated.

Exact probability tables

A large amount of computing resources is required to compute exact probabilities for the Kruskal-Wallis test. Existing software only provides exact probabilities for sample sizes less than about 30 participants. These software programs rely on asymptotic approximation for larger sample sizes. Exact probability values for larger sample sizes are available. Hollander and Wolf (2003) published exact probability tables for samples as large as 45 participants. Meyer and Seaman (2006) produced exact probability distributions for samples as large as 105 participants.

3.4.2 The One – Way Anova Assumptions

a) All the units must be homogenous (similar).

b) Experimental errors are normally distributed within each treatment.

c) The population have equal variance.

Advantages

a) It is easy to design since the allocation of treatments to unit is completely at random.

b) There is no need for equal replication. Therefore, there is no need for estimating missing value.

c) Analysis is very simple and there is high degree of freedom for residuals indicating high precision.

Disadvantage

a) The design is inefficient if the units are non homogenous.

Table 3.1: The Anova Table for One- Way

Source of Variation	Degree of Freedom	Sum of squares	Mean square	F
Treatment	t-1	Sum of square Treatment	Mean square Treatment	MST/MSE
Error	N-t	Sum of square Error	Mean square Error	
Total	N-1	Sum of square Total		

The total sum of squares is given by:

$$SST = \sum_{i=1}^{r}\sum_{j=1}^{t} y_{ij}^{\ 2} - \frac{y_{..}^{\ 2}}{N}$$

The treatment sum of squares is given by:

$$SST_r = \frac{1}{n}\sum_{i=1}^{r} y_{i.}^{\ 2} - \frac{y_{..}^{\ 2}}{N}$$

The error sum of square is given by

SSE= SST -SSTr

Reject the null hypothesis if F- calculated > F-tabulated

The F- calculated follows central F- distribution when H is true, otherwise the F-calculated is non central F- distributed with non centrality parameter lambda (λ).

That is, under

H_0, $F \sim F_{v1,v2}$

Under

H_1, $F \sim F_{v1,v2}$

Where: v_1 and v_2 are the numerator and denominator degrees of freedom respectively.

3.5 SIMULATION

Simulation is a way to model random events, such that simulated outcomes closely match real-world outcomes. By observing simulated outcomes, researchers gain insight on the real world.

HOW TO CONDUCT A SIMULATION

A Simulation is useful only if it closely mirrors real-world outcomes. The steps required to produce a useful Simulation are presented below;

i. Describe the possible outcomes.
ii. Link each outcome to one or more random numbers.
iii. Choose a source of random numbers.
iv. Choose a random number.
v. Based on the random number, note the simulated outcome.
vi. Repeat steps 5 and6 multiple times; preferably, until the outcomes shows a stable pattern.

vii. Analyze the simulated outcomes and report results.

CHAPTER FOUR

PRESENTATION OF RESULTS AND DISCUSION

4.1 Introduction

This chapter presents the result and discussion of the analysis conducted for the research study.

4.2 Test of Hypotheses and Region of Decision under normal and exponential distribution for $n < 30$, $n = 30$ and $n > 30$

1. H_{01}: the percentage MGD difference for the two different speed rate of the threshing machine is equal to zero at sample size of $n < 30$, $n = 30$ and $n > 30$.

   H_{11}: the percentage MGD difference for the two different speed rate of the threshing machine is not equal to zero at sample size of $n < 30$, $n = 30$ and $n > 30$.

2. H_{02}: the mean (median) percentages of MGD for the three speed rate of the threshing machine are equal at sample size of $n < 30$, $n = 30$ and $n > 30$.

   H_{12}: the mean (median) percentages of MGD for the three speed rate of the threshing machine are not equal at sample size of $n < 30$, $n = 30$ and $n > 30$

The null hypotheses will be rejected if the test statistic probability value (P-value) is less than alpha $\alpha = 0.05$, 0.01 and 0.1 significance level. Note, in the second hypothesis the mean is for the parametric test while the median in bracket is for the nonparametric test.

4.3 Analysis for; *n*<30, *n*=30 and *n*>30 of the Percentage of Mechanical Grain Damage (MGD) for Independent T-test and Wilcoxox Rank Sum Test

Table 4.1: Number of rejection out of 50 trials under normal distribution for two independent samples

	α = 0.01		α = 0.05		α = 0.1	
Sample size	T	U	T	U	T	U
45	1	1	2	2	2	2
30	1	2	3	3	3	4
10	2	2	3	3	6	6

From the table above the power (1 - β) for each significant level are:

For T- test, at ∝= 0.01 P (rejection) = P (H_0^I/ n=45) = $^1/_{50}$

P (rejection) = P (H_0^I/ n = 30) = $^1/_{50}$

P (rejection) = P (H_0^I/ n = 10) = $^2/_{50}$

∴ Overall P(rejection) = P(H_0^I/ $\forall n$) = $^4/_{150}$

β = P (rejection / H_1)

$$\beta = \frac{P(H^{l}_{0\forall n})}{P(H^{l}_{0}/n=45)+P(H^{l}_{0}/n=30)+P(H^{l}_{0}/n=10)}$$

β = 0.333375

Power $(1-\beta)$ = 1 – 0.333375 = 0.6666

For Wilcoxon rank test at α = 0.01

= P (H_0^I/ n=45) = $\frac{1}{50}$

= P (H_0^I/ n=30) = $\frac{2}{50}$

= P (H_0^I/n=10) = $\frac{2}{50}$

Overall P(rejection) = P(H_0^I/ $\forall n$) = $\frac{5}{150}$

∴ β = 0.3333

And Power = 1 – 0.3333 = 0.6667

For T- test at α = 0.05

= P (H_0^I/ n=45) = $\frac{2}{50}$

= P (H_0^I/ n=30) = $\frac{3}{50}$

= P (H_0^I/n=10) = $\frac{3}{50}$

Overall P(rejection) = P(H_0^I/ $\forall n$) = $\frac{8}{150}$

$\therefore$ β = 0.3333

And power = 1 – 0.3333 = 0.6667

For Wilcoxon test at 0.05

= P (H_0^I/ n=45) = $\frac{2}{50}$

= P (H_0^I/ n=30) = $\frac{3}{50}$

= P (H_0^I/n=10) = $\frac{3}{50}$

Overall P(rejection) = P(H_0^I/ $\forall n$) = $\frac{8}{150}$

$\therefore$ β = 0.3333

And Power = 1 – 0.3333 = 0.6667

For T–test at α = 0.1

= P (H_0^I/ n=45) = $\frac{2}{50}$

= P (H_0^I/ n=30) = $\frac{3}{50}$

= P (H_0^I/n=10) = $\frac{6}{50}$

Overall P(rejection) = P(H_0^I/ $\forall n$) = $\frac{11}{150}$

$\therefore$ β = 0.3333

And Power = 1 – 0.3333 = 0.6667

For Wilcoxon test at 0.1

β = 0.3333

$\therefore$ Power = 1 – β = 0.6666

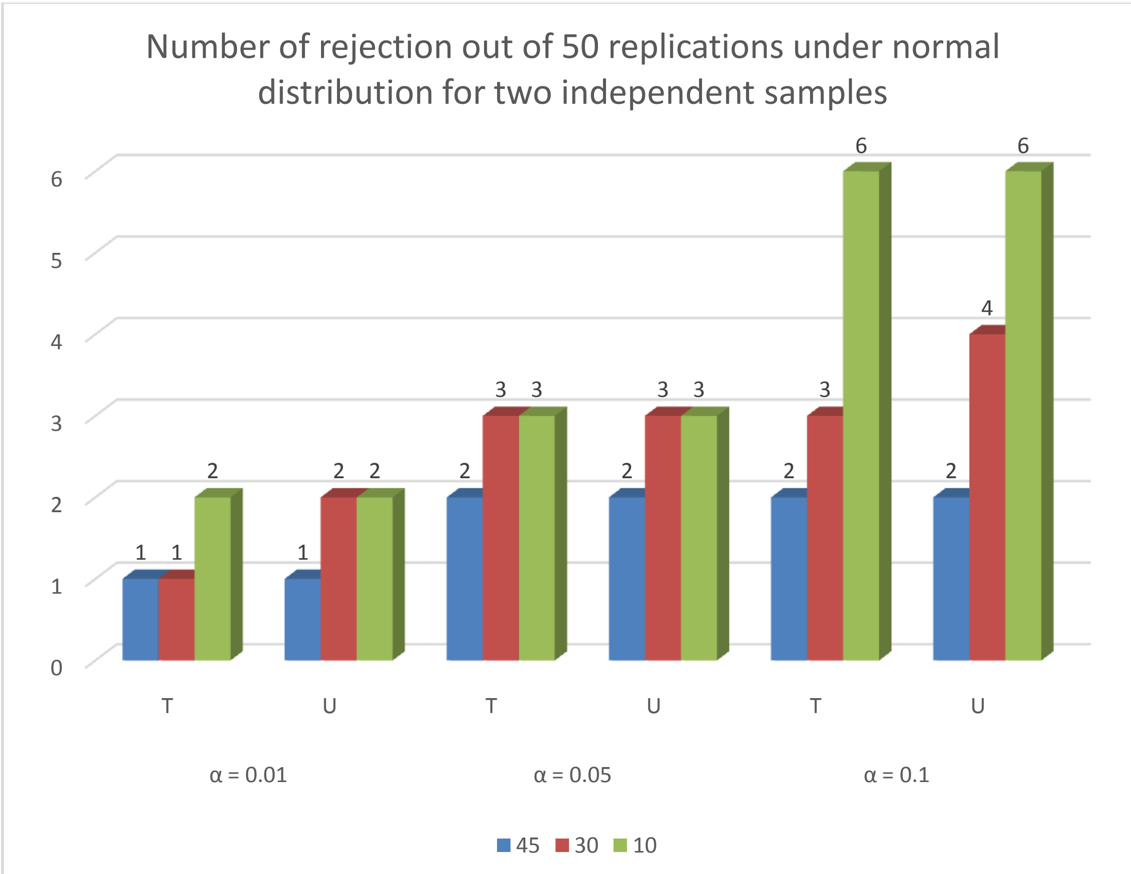

Fig 4.1: Bar Chart for Number of Rejections

From Table4.1 above where the data is simulated fifty (50) times under normal distribution, the independent sample t-test analysis was performed and compared with its equivalent test, the Wilcoxon rank-sum test at $\propto$ = 0.05, 0.01 and 0.1. It was observed from the analysis performed that at n = 10 and n = 45 the power of the test are the same that is the two tests performed equally at all levels of α but at n=30 the tests perfomed equally at α = 0.05 but at α= 0.01 and 0.1 the nonparametric test is as powerful as the parametric test given adequate sample size. Hence, from these analyses, the issue of power between nonparametric and parametric test would not be of much concern as far as a sample size of at least 10 is taken. This gives an eminent advantage of the nonparametric methods because they are easier to perform and no underlying assumptions about the mother population.

And in Fig 4.1, it is observed that as the level of significance α increases, the probability of Type 11 error β decreases (an inverse relationship) hence the power (1-β) increases as the sample size decreases.

Table 4.2: Number of rejection out of 50 replications under exponential distribution for two independent samples (independent T-test and Wilcoxox Rank-Sum Test)

	α = 0.01		α = 0.05		α = 0.1	
Sample size	T	U	T	U	T	U
45	1	0	4	3	6	5
30	0	1	2	1	5	4
10	0	0	0	1	1	3

From the table above the power (1 - β) for each significant level are;

For T-test at α = 0.01

= P (H_0^I/ n=45) = $\frac{1}{50}$

= P (H_0^I/ n=30) = $\frac{0}{50}$

= P (H_0^I/n=10) = $\frac{0}{50}$

Overall P(rejection) = P($H_0^I / \forall n$) = $\frac{1}{150}$

∴ β = 0.3333

And power = 1 – 0.3333 = 0.6667

ForWilcoxon test at $\propto$ =0.01

= P (H_0^I/ n=45) = $\frac{0}{50}$

= P (H_0^I/ n=30) = $\frac{1}{50}$

= P (H_0^I/n=10) = $\frac{1}{50}$

Overall P(rejection) = P(H_0^I/ $\forall n$) = $\frac{1}{150}$

$\therefore \beta$ = 0.3333

And Power = 1 – 0.3333 = 0.6667

For T-test at $\propto$= 0.05

= P (H_0^I/ n=45) = $\frac{4}{50}$

= P (H_0^I/ n=30) = $\frac{2}{50}$

= P (H_0^I/n=10) = $\frac{0}{50}$

Overall P(rejection) = P(H_0^I/ $\forall n$) = $\frac{6}{150}$

$\therefore \beta$ = 0.3333

And Power = 1 – 0.3333 = 0.6667

For Wilcoxon test at $\propto$ = 0.05

= P (H_0^I/ n=45) = $\frac{3}{50}$

= P (H_0^I/ n=30) = $\frac{1}{50}$

= P (H_0^I/n=10) = $\frac{1}{50}$

Overall P (rejection) = P(H_0^I/ $\forall n$) = $\frac{5}{150}$

$\therefore \beta$ = 0.3333

And Power = 1 – 0.3333 = 0.6667

For T-test $\propto$=0.1

$\therefore \beta$ = 0.3333

And Power = 1 – 0.3333 = 0.6667

For Wilcoxon test at $\propto$ = 0.1

$\therefore \beta$ = 0.33337

And Power = 1 – 0.3333 = 0.6667

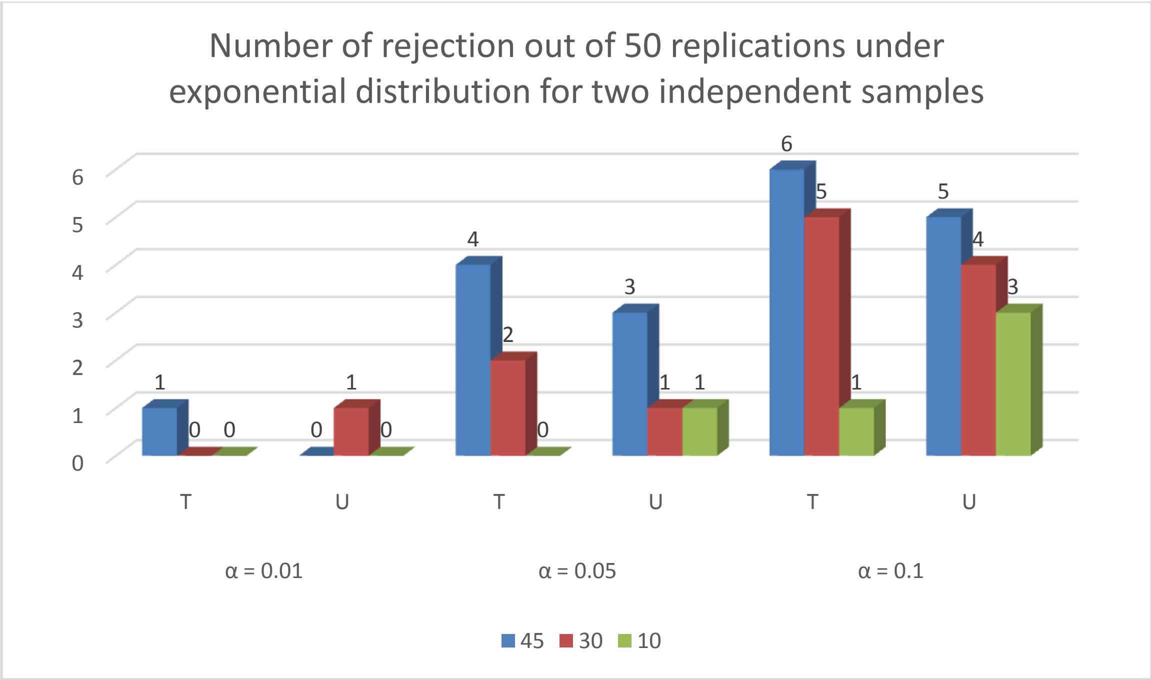

Fig 4.2 Bar Chart for Number Rejections

From Table4.2 above where the data is simulated fifty (50) times under the exponential distribution, also the independent sample t-test analysis was performed and compared with its equivalent, the Wilcoxon rank-sum test at $\propto$ = 0.05, 0.01 and 0.1, for n=45 and also at $\propto$ = 0.05 and 0.1 for n= 30 it was observed that the parametric test is more powerful. But the nonparametric is as powerful as the parametric test for n= 30 and n=10, at α = 0.01 and $\propto$ = 0.05 &0.1 respectively given. Hence, from these analyses, the issue of power between nonparametric and parametric test would not be of much concern as far as a sample size of at least 10 is taken. This gives an eminent advantage of the nonparametric methods because they are easier to perform and no underlying assumptions about the mother population.

And in Fig4.2, it is observed that as the level of significance α increases, the probability of Type 11 error (β) decreases (an inverse relationship) hence the power (1-β) increases as the sample size increases.

4.4 Analysis for; *n*<30, *n*=30 and *n*>30 of the Percentage of Mechanical Grain Damage (MGD) for one – way Anova and Kruskal Wallis Test

Table4.3: Number of rejection out of 50 replications under normal distribution for more than two independent samples

	α = 0.01		α = 0.05		α = 0.10	
Sample size	F	H	F	H	F	H
45	2	0	3	2	4	3
30	1	2	3	2	3	3
10	2	1	6	3	9	6

From the table above the power(1 - β) for each significant level are;

For One–Way Anova at $\propto$=0.01

$= \text{P } (H_0^I/ \text{ n=45}) = \frac{2}{50}$

$= \text{P } (H_0^I/ \text{ n=30}) = \frac{1}{50}$

$= \text{P } (H_0^I/\text{n=10}) = \frac{2}{50}$

Overall P (rejection) = $\text{P}(H_0^I/ \ \forall n) = \frac{5}{150}$

$\therefore \beta = 0.3333$

And power = 1 – 0.3333 = 0.6667

For Kruskal-Wallis at $\propto$= 0.01

$= \text{P } (H_0^I/ \text{ n=45}) = \frac{0}{50}$

$= \text{P } (H_0^I/ \text{ n=30}) = \frac{2}{50}$

$= \text{P } (H_0^I/\text{n=10}) = \frac{1}{50}$

Overall P (rejection) = $\text{P}(H_0^I/ \ \forall n) = \frac{5}{150}$

$\therefore \beta = 0.3333$

And power = 1 – 0.3333 = 0.6667

For One-Way Anova at $\propto = 0.05$

$\therefore \beta = 0.3333$

And power = 1 – 0.3333 = 0.6667

For Kruskal-Wallis at $\propto = 0.05$

$\beta = 0.3333$

Power = $1 - \beta = 06666$

for One-Way Anova at $\propto = 0.1$

$\beta = 0.3333$

$\therefore$ Power = $1 - \beta = 0.6666$

For Kruskal-Wallis at $\propto = 0.1$

$\beta = 0.3333$

Power = $1 - \beta = 06666$

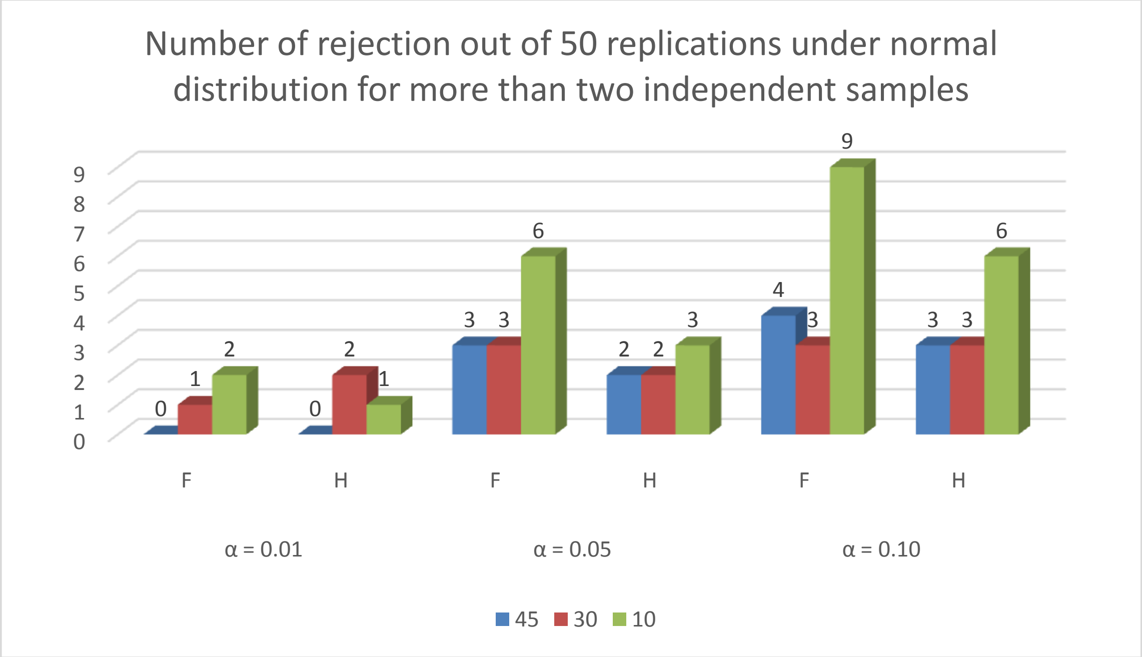

Fig 4.3 Bar Chart for Number of Rejection

From Table4.3 above where the data is simulated fifty (50) times under the normal distrbution, the one - way Anova was performed and compared with its equivalent, Kruskal-Wallis test, for the three sample sizes at α= 0.05 and 0.1 and also at α= 0.01 for *n*=45 and n=10 the parametric test is more powerful execpt for *n*=30 the nonparametric is as powerful as the parametric test given adequate sample size. Hence, from these analyses, the issue of power between nonparametric and parametric test would not be of much concern as far as a sample size of at least 10 is taken. This gives an eminent advantage of the nonparametric methods because they are easier to perform and no underlying assumptions about the mother population.

And in Fig4.3, it is observed that as the level of significance α increases, the probability of Type 11 error β decreases (an inverse relationship) hence the power (1-β) increases as the sample size decreases.

Table4.4: Number of rejection out of 50 replications under exponential distribution for more than two independent samples

	α = 0.01		α = 0.05		α = 0.10	
Sample size	F	H	F	H	F	H
45	2	0	4	2	7	9
30	1	1	5	2	7	5
10	0	0	1	4	4	5

From the table above the power (1 - β) for each significant level are

For One-Way Anova at ∝= 0.01

β = 0.3333

∴ Power (1 - β) = 0.6666

For Kruskal-Wallis at ∝=0.01

β = 0.3333

$\therefore$ Power $(1 - \beta) = 0.6666$

For One-Way Anova at $\propto = 0.05$

$\beta = 0.3333$

$\therefore$ Power $(1 - \beta) = 0.6666$

For Kruskal-Wallis at $\propto = 0.05$

$\beta = 0.3333$

$\therefore$ Power $= (1 - \beta) = 0.6666$

For One -Way at $\propto = 0.1$

$\beta = 0.3333$

$\therefore$ Power $(1 - \beta) = 0.6666$

For Kruskal-Wallis at $\propto = 0.1$

$\beta = 0.3333$

$\therefore$ Power $(1 - \beta) = 0.6666$

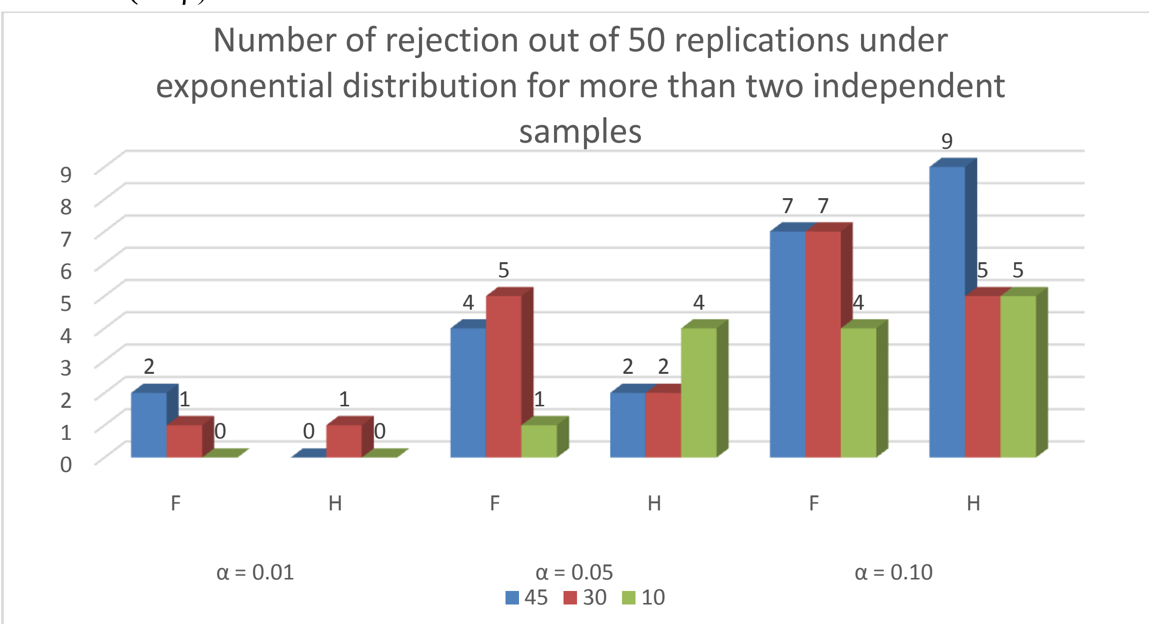

Fig4.4 Bar Chart for Number of Rejection

From Table4.4 above where the data is simulated fifty (50) times under the exponential distrbution, the one - way Anova was performed and compared with its equivalent, Kruskal Wallis test, for n=45, n=45 & 30 and n=45&10 at α= 0.01, 0.5 and 0.1

respectively the parametric test is more powerful while for n= 10 and n=45&10 at α = 0.05 and 0.1 respectivly the nonparametric is as powerful as the counterpart and lastly for n=30&10 at α= 0,01 they performed equally. Hence, from these analyses, the issue of power between nonparametric and parametric test would not be of much concern as far as a sample size of at least 10 is taken. This gives an eminent advantage of the nonparametric methods because they are easier to perform and no underlying assumptions about the mother population.

And in Fig4.4 it is observed that as the level of significance α increases, the probability of Type II error β decreases (an inverse relationship) hence the power (1-β) increases as the sample size increases.

Therefore the general conculsion is that the power for both parametric and non-parametric test under the normal and exponential distribution for the three significant level are almost the same. This signify that the non-parametric (Wilcoxon test and Kruskal Wallis test) are so powerful as the parametric (T-test and One-Way Anova) given adequate sample size. This shows that the nonparametric could give as much power as the parametric method while still with holding the advantage of conformance with normality assumptions and flexibility.

CHAPTER FIVE

SUMMARY, CONCLUSION AND RECOMMENDATION

5.1 Introduction

In this chapter, we present the summary, conclusion and recommendations based on the results obtained in chapter four as an overview of the study.

5.2 Summary and Conclusion

We recall that the main objectives of this research work is to assess the parametric test strength with its nonparametric equivalents. The comparison of the class of test were subject to three cases which are the sample size, $n < 30$, n=30 and $n > 30$. From, the analysis performed for Two Independent Samples and for more than Two Independent Samples under the normal distribution it was observed that the small sample sizes as higher powers and under the exponential distribution the large sample sizes have higher powers.

As stated by Siegel (1956), Runyon (1977) and Goon and Gupta (2003) that a particular test (A) may be more powerful than its counterpart test (B) when their *n*'s are equal but its counterpart (B) may be as powerful as its efficiency (A) if its sample size is increased. This result could be largely influenced by outliers in the data set and as it is well known the parametric test make use of the mean which is highly affected by outliers compared to the non-parametric test that make use of the median which is highly resistant to outliers Masoud and Rahim, (2010) likewise, the mean square error, MSE for the ANOVA F-test can be unsecured, Sergio and Joan (2001) stated that like variance, MSE has the disadvantage of heavily weighting outliers. They said it is because of the squaring of each term, which effectively weights large errors more heavily than small ones. This property, undesirable in many applications, has led researchers to use alternatives such as the mean absolute error, or those based on the median.

However, when the samples are continuous but the experiment is not sure of the assumption of normality and equllity of variance, the nonparametric is to be considered. Hence, it can be concluded from this research that the parametric and non-parametric test are reliable when their assumptions are met and their efficiency can be greatly determined by their sample size therefore, where the sample size is not small, any of the two approaches will perform well for the purpose of such comparison irrespective of the distribution.

5.3 Recommendation

In this research study we present the following recommendation when trying to decide on which test to apply in any study.

1. We recommend that the assumption criteria must be met by fully exploring the various test to ascertain the criteria present and this assumptions are; independency, normality, constant variance and linearity. If these conditions are met then the parametric test is preferred to the nonparametric test.
2. Test for the presence of outliers should be done and decision on whether to delete the data point(s) or not and if it is actually an important data point(s) that have to be included then care should be taken on using the parametric test or better still test that make use of the median (nonparametric test) should be used.

3. If the data collected is in interval or ratio scale the parametric test should be applied except the underlying conditions are not met even after data transformation because it could lead to loss of in-depth analytical information but if the data is ordinal or nominal all temptations towards the parametric test should be abolished.

4. Also when samples are continuous but the experiment is not sure whether the assumption of normality and homogeneity of variance are valid, the nonparametric is to be considered.

5. We recommend this research for further open criticism and improvement to researchers in the field of statistics and related fields.

5.4 Contribution to Knowledge

The old perception of the nonparametric being less powerful in all cases has been proven to be otherwise because this referral work has shown that nonparametric can be as powerful as parametric given sufficient sample size.

The work has also shown that as the sample size increases, the power also increases for exponential distribution regardless of the level of significance. While for the normal distribution, the level of significance influences the relationship between the sample size and the power of test.

5.5 Recommendation for Further Studies.

For further studies it is recommended that other parametric and nonparametric distributions should be compared using small and large samples size under different distributions and also for three or more independent groups.

REFERENCES

Abebe, A. (2009): Introduction to Design and Analysis of Experiment with the SAS System(Stat 7010 Lecture Notes).Department of Discrete and Statistical Sciences, Auburn University, Auburn.Auburn University Press.

Andrew, M. J, Nigel, R. and Silvana, R (2011): A Comparison of Parametric and Nonparametric, University of York, York. Ac.

Atman, D. G. (1991): Practical Statistics for Medical Research London; Chapman and Hall (monograph).

Atman, D. G. (1948): Parametric and Non-parametric Methods,3^{rd} ed Curt Hinrichs.

Bradley, E. (1996): Empirical Bayes Methods for combining likelihoods J. Amer statistics Association.

Breheny, P. (2012): Connection between Parametric and Nonparametric Theory, The Oxford handbook Applied Nonparametric and Semi-parametric Econometrics course. Hero Inc.

Bridge, P. D. and Sawilowsky, S. S.(1999): Increasing Physicians Awareness of the Impact of Statistics on Research Outcomes: Comparative Power of the t-test and Wilcoxon Rank-Sum test, Journal of Clinical Epidemiology.

Briggs, A. H, Wonderling, D. E, and Mooney, C. Z(1997): Pulling Cost-Effectiveness Analysis up by its boot steps, A Nonparametric Approach to Confidence Interval Estimation. Health Econ.

Cambell, G. K(2006): An Introduction to Nonparametric Statistics for Health Scientists. University of Albert Health Journal, Volume 3. Issue 1.20 reviews

Cohen, J. (1988): Statistical Power Analysis for the Behavioural Sciences, 2nd ed. Hillsdale, NJ: L. Erlbaum Associates.

Conover, W.J.(1999): Practical Nonparametric Statistics, 3^{rd} ed John Wiley and sons, New York.

Cyprain A. O.(1990): An Introduction to Applied Statistical Methods in the Science, Nobern Avocation Publishing Company Enugu. PP. 378

Davis, C. S.(1991): Semi Methods in the Science- Parametric and Non-parametric Methods for the Analysis of Repeated Measurements with Application to Clinical Trials. Statistics in Medicine, 10(12); 1959-1980.

Fisher L. D and Van Belle G.(1993): Biostatistics; A Methodology for the Health Science, New York.

Friedman M. (1937): The use of Ranks to avoid the Assumption of Normality in the Analysis of Variance. Journal of the American Statistical Association.

Goon, A. M. and Gupta, M. K.(2003): An Outline of Statistical Theory, The World Press Private Limited. PP. 449 – 489.

Greene, W. H. (2000): Econometric Analysis, 4th edition. Upper Saddle River, NJ: Prentice Hall.

Hinton P.R (2004): Statistics Explained, 2nd edition, New York: Routledge.

Hollander, M. and Wolfe, D. A.(1973): Nonparametric Statistical Method, New York.

Hotelling, H. and Pabst, M.(1936): Rank Correlation and Test of significance involving no assumptions of Normality. Annuals for the Mathematical Statistician.

Kendall, M. G(1938): A new Parametric and Nonparametric measure of Rank Correlation Biometrika.

Masoud, Y. and Rahim, M.(2010): The Effect of Outliers on Robust and Resistant Coefficient of Determination in the Linear Regression Models, International Journal of Academic Research.

Meyer, J.P. And Seaman .M.A.(2014): A Comparison of the Kruskal-Wallis Distribution to Asymptotic Approximation for all Sample Size up to 105.Journal of Experimental Education. 81(2)

Montgomery, D. C.(2005): Design and Analysis of Experiments, 5th edition, John Wiley and Sons New York.

Peter, C. and Brien, O.(1984): Procedures for Comparing samples with Multiple endpoints. International Biometric society,40(4).

Pett, M.A. (1992): Nonparametric Statistics in Health Care Research: Statistics for Small Sample and Unusual Distribution.2^{nd} Edition. SAGE Publication, Inc.

Pitman E. (1948): (Pitman Estimation) Lecture Note on Nonparametric Statistical Inference Columbia univ., NY.

Runyon, R. P.(1977): Nonparametric Statistics a Contemporary Approach, Addison Wesley Publishing Company, Inc. Philippines Copy right.

Sawilowsky, S. and Blair, R. C.(1992): A more realistic look at the robustness and Type II error properties of the T-test to department from population normality, Psychological Bulletin.

Sergio, B. and Joan, C.(2001): Oriented Principal Component Analysis for Large margin Classifiers, Neural Networks 14(10).

Siegel S.(1956): Nonparametric Statistics for the Behavioural Sciences, New York: McGraw-Hill.

Siegel, S. and Castellan, N. J.(1988): Nonparametric Statistics for the Behavioural Sciences, 2nd ed. New York: McGraw Hill.

Simon P.(2008): Comparison of Parametric and Nonparametric References Data in Motor Unit Potential Analysis, Muscle and Nerve, Wiley Periodical,Inc.

Smirrov, K.(1936): Power Comparison of Shapivo Wilk, K.S, Lilliefor and Anderson-Darling Test. Journal of Statistical Modelling and Analytics

Whitehead, A. And Whitehead, J.(1996): A General Parametric Approach to the met Analysis of Randomized Clinical Trials, Economic and Health Statistics, John Wiley and Sons Inc.

Vickers, A. J. (2005): Parametric versus non-parametric Statistics in the Analysis of Randomized Trials with Non-normally Distributed Data BMC Medical Research Methodology Volume: 5, Issue: 1, BioMed Central.

Zimmerman, D. W. and Zumbo, B. D.(1990**):** The Effect of Outliers on the Relative Power of Parametric and Nonparametric Statistical Tests**,** Perceptual and Motor Skills, 71.