POWER COMPARISON OF SOME PARAMETRIC AND NON-PARAMETRIC TESTS

[1]BASHIR ALHAJI MUSTAPHA, [2]ALHAJI MODU ISA, AND [3]TIJJANI ADAMU

[1]Department of Statistics, The Federal Polytechnic Damaturu, Yobe State, Nigeria. [2]Department of Mathematics and Computer Science, Borno State University, Nigeria. [3]Department of Statistics, University of Maiduguri, Borno State, Nigeria.
bashirmustapha65@gmail.com

ABSTRACT

In this study power of statistical tests is used as a criterion for selection of best method. Experiments were subjected to both parametric and non-parametric methods of analysis; large and small samples were considered in each case. It was found that for large samples, the parametric tests have an average power of 80.57% and the non-parametric has 77.74% for small samples, the parametric tests gave an average power of 52.43% while non-parametric tests recorded 35.76%. When both large and small samples were combined, the parametric tests gave an average power of 66.5% while the non-parametric tests recorded 56.85%. It could be concluded that parametric tests are generally more powerful than their non-parametric counterparts and also large samples are more powerful than small samples.

Keywords: Power, Parametric, Non-Parametric, Comparison, Test

INTRODUCTION

The new encyclopedia Britannica (1998) gave a concise definition of statistics as ("the science of collecting, analyzing, presenting and interpreting data"). On that basis, therefore, a statistician is basically concerned with the chance outcomes that occur in scientific investigation Tarrel (2021). He may be interested in the number of accidents occurring at a certain location, the outcome when a coin is tossed, or the amount of residue deposited in a chemical experiment.

Stapor and Stapor (2020) claimed that in statistical inference we are concerned in making decisions and drawing conclusions about a large set of data which we only have partial knowledge based on a smaller subset of the data; and so statistics provides tests which formalize and standardize our procedures for drawing conclusions. Therefore, one can say without any loss of generality that inferential statistics deals with two types of problems; estimation of population parameters and tests of hypotheses.

A common problem for statistical inference is to determine, in terms of probability, whether observed inferences between two samples signify that the population samples are themselves really different Tintle et al., (2020). It is clear that whenever we collect two groups of scores by random methods we are likely to find that the scores differ to some extent. And differences occur on the basis of operations of chance. Then the question here is how can we determine in any given samples whether the observed differences are merely due to chance or not? Tredennick *et al.,* (2021) claimed that the procedures of statistical inference enables us to determine in terms of probability, whether the observed difference is within the

range which could easily occur by chance or whether it is so large that it signifies that the two samples are probably from two different populations.

Cann *et al.,*(2004) is of the view that in the development of modern statistical methods, there emerged a technique of inference which made a good many assumptions about the nature of the population from which the scores were drawn. Since populations values are parameters these statistical techniques are called parametric.

Also of recent, we have witnessed developments of a large number of inferential methods, which do not make numerous or stringent assumptions about parameters. These new techniques are called distribution-free or non-parametric tests, and require fewer assumptions Yao et al.,(2021).

It was reported by many authors such as Lee-Yaw et al., (2022), Davies et al., (2020) and Dalmaijer et al., (2022), that the power of a test is its sensitivity and ability to detect relationship that exists on the population. It is the probability of rejecting the null hypothesis when it is false. Therefore a good test should have a very low β and high percentage of power.

A hypothesis is an assumption or speculation we make about a population parameter Lan et al.,(2020). In hypothesis testing, we must state the assumed or hypothesized value of the population parameter before we begin sampling. The assumption we wish to test is called the null hypothesis and is symbolized H_0 Trafimow *et al.,* (2021). If our sample results fail to support the null hypothesis, we must conclude that something else is true. Whenever we reject the null hypothesis, the conclusion we do accept is called the alternative hypothesis and is symbolized H_1.

Maier and Lakens (2022) claimed that when decision is to be made about differences, then Ho should be tested against H_1. H_1 constitutes the assertion that is accepted if Ho is rejected. Also Maxwell and Delaney (1990) believed that the nature of the research hypothesis determines how H_1 should be stated. If the research hypothesis simply states that two groups will differ with respect to means, then H_1 is that $\mu_0 \neq \mu_1$ which is called a one-tailed test. But if the theory predicts the direction of the difference i.e., that one specified group will have a larger mean than the other, then H_1 may be either that $\mu_1 > \mu_2$ or $\mu_1 < \mu_2$ in which case is called a two tailed test. The two tailed test is the most frequently used test.

Scheel *et al.,* (2021) believed that the hypothesis testing is on the basis of certain percentage level of significance. Therefore, assuming the hypothesis is correct, the significance level indicates the percentage of sample means that is outside certain limits.

Wasserman et al., (2020) observed that there is no single standard or universal level of significance for testing hypothesis. It is possible to test a hypothesis at any level of significance. But remember that the choice of the minimum standard for an acceptable probability, or significance level, is also the risk we assume of rejecting a null hypothesis when it is true. Therefore, the higher the significance level used for testing a hypothesis, the higher the probability of rejecting a null hypothesis when it is true (Di Leo & Sardanelli 2020).

Tredennick *et al.,* (2021) identified three different procedures for testing hypothesis with all of them leading to the same decisions when the same probability (and risk) standards are used. The first approach that was developed is the critical value approach to hypothesis testing. Here the so called critical values of the test statistic that would dictate rejection of a hypothesis are determined, and then the observed test statistic is compared to the critical values.

The second approach which is more recently developed is the P-value approach. It has become more popular because it is the one most easily applied with computer software. This approach is based on determining the conditional probability that the observed value of a sample of statistic could occur by chance, given that a particular claim for the value of the associated population parameter is in fact correct. Finally, it was viewed by Serdar *et al.,* (2021) that the critical issue in statistical conclusion validity is power. Therefore in the opinion of Noble et al., (2021) and Joshi and Dhakal (2021), the power of a test is its sensitivity or ability to detect relationships that exist on the population, and so it is the complement of a type II error. In conventional terms, power is the probability of rejecting the null hypothesis when it is false and equal 1 minus the probability of a type II error $(1-\beta)$.

The power of a statistical analysis is partly a function of the statistical test employed in the analysis Dalmaijer *et al.,* (2022). Smith (2020) put it this way, that, a statistical test is a good one if it has a small probability of rejecting H_o when H_o is false. In selecting a statistical test we should simply select the one that has the larger probability of rejecting Ho when it is false.

Aims and Objectives of the Study

The aims and the objectives of this study are:

- Consider small and large samples analyze the experiments by appropriate parametric test and also by the corresponding non-parametric test.
- Use the criterion of power to determine which of the two methods is more powerful.

LITERATURE REVIEW

In trying to give a mathematical concept of the terms samples and populations, Karunasingha (2022) claimed that the terms can be described by using measures such as the mean, median, mode and standard deviation, when these terms describe the characteristics of a sample, they are called statistics. When they describe the characters of a population, they are called parameters. In brief therefore, "a statistic is a characteristic of a sample and a parameter is a characteristic of a population".

Luengo *et al.,* (2020) ascertain that, the methods of inference we usually use are called parametric methods. In his words "parameters are indices, they index (or label) individual distributions within a particular family" To round it up, Joshi et al., (2020), Smith (2020), Karunasingha (2022) and Luengo et al., (2020) described parametric methods as those that involve estimating or testing the values of parameter(s), which are usually, population means or proportions. In a parametric test a sample statistic is obtained to estimate the population parameters. However, because this estimation process involves a sample, a sampling distribution, and a population, certain parametric assumptions are required to ensure that all components are compatible with each other.

Smith (2020) gave out conditions, which must be satisfied before using any parametric test, the conditions includes;

1. The observations must be independent.
2. The observations must be drawn from normally distributed population.
3. These populations must have the same variances.
4. The variables involved must have been measured in at least an interval scale.

A scale of measurement is a particular way of assigning numbers or symbols to measure something Dalmaijer *et al.,* (2022). It is believed that all forms of statistical tests rests on the scale of measurement. Therefore measurement theory is important because it helps us to avoid making meaningless statements.

The Unpaired t-test

The unpaired *t*-test does not require that the two groups be paired in any way, or even of equal sizes Havre *et al,* (2002). Here, we wish to know if the differences between the groups are real (statistically significant) or could have arisen by chance.

The calculations involved in an unpaired t-test are slightly more complicated than for the paired test Townsend *et al.,* (1999).

Cann (2004) gave out steps for performing the unpaired *t*-test. This is as follows:

1. Plot a histogram of the data sets to confirm that they are normally distributed. If not stop and use a non-parametric method.
2. Start with the hypothesis (H_o), there is no difference between the populations of measurements from which samples have been drawn. (H_A; there is a difference).
3. Set a value for α (significance level).
4. Check the variance of each group; if the variances are very different, you can reject the hypothesis that the samples are from a common population, even though their means might be similar. However, it is possible to go on and perform a *t*-test.
5. Calculate *t* using the formula:

$$t = \frac{\overline{X}A - \overline{X}B}{\sqrt{\left(SE_A\right)^2 + \left(SE_B\right)^2}}$$

Where $\overline{X}$ = means of groups *A* and *B*, respectively and $SE = \frac{SD}{\sqrt{N}}$

6. Look up the value of t for the correct number of degrees of freedom and for a one or two tailed test.

$df = \left(n_A + n_B\right) - 2$ where $n_A \mid B$ = the number of values in the groups compared. Remember that the sign of $t\left(+/-\right)$ does not matter. Assume that *t* is positive.

7. If the calculated value of *t* is greater than the critical value, H_o is rejected, i.e. there is evidence of a statistically significant difference between the groups.

If the calculated value of *t* is less than the critical value, H_o is accepted, i.e. there is no evidence of a statistically significant difference between the groups.

One-Way Analysis of Variance

The one-way ANOVA Tests differences in single interval dependent variable among two, three or more groups formed by the categories of a single categorical independent variable. It is also known as univariate ANOVA, single ANOVA, simple classification ANOVA, or one factor ANOVA, this design deals with one independent variable and one dependent variable Tabachnick & Fidell, (2007).

In terms of statistical linear model (Montgomery, 1984) described the observation of *K* treatments or different levels of a single factor that we wish to compare as

$$X_{ij} = \mu + \alpha_i + \varepsilon_{ij}, \begin{cases} i = 1,2,...,k \\ j = 1,2,...,n \end{cases}$$

Where X_{ij} is the (*ij*)th observation, μ is a parameter common to all treatments called overall mean, α_i is a parameter unique to the *i*th treatment called *i*th treatment effect, and ε_{ij} is a random error component. Our objectives will be to test appropriate hypothesis about the treatment affects and to estimate them.

The statistical model above described two different situations with respect to the treatment effects. This could be fixed-effects model or random effects model, depending on the experimenter. Therefore one-way ANOVA could be designed as either fixed- effects or random effects.

The table below is the summary table for the one factor completely randomized design of the analysis of variance, including all computational formulas. Because of the need to use a system, which can be extended logically to two-way analysis of variance, we use similar symbols.

Table1: Summary table for one-way ANOVA (Treatment groups need not be equal), (Kazmier, 1996, 229)

Source of variation	Degrees of freedom (df)	Sum of squares (ss)	Mean square	F ratio
Among treatment groups (A)	$k-1$	$SSA = \sum \frac{T^2}{n_k} - \frac{T^2}{N}$	$MSA = \frac{SSA}{k-1}$	$F = \frac{MSA}{MSE}$
Sampling error (E)	$N-k$	$SSE = SST - SSA$	$MSE = \frac{SSE}{N-k}$	
Total (T)	$N-1$	$SST = \sum_{i=1}^{n}\sum_{n=1}^{k} X^2 - \frac{T^2}{N}$		

where *Tk*= the sum (total) of the values in a particular treatment group

T = the sum of the sampled values in all groups combined

N = total number of subjects

k = the number of groups.

The general form of the null hypothesis in the analysis of variance makes reference to the relevant component of the linear model. Thus, for the one-way analysis of variance the null and alternative hypothesis can be stated as

H_o: $\mu_1 = \mu_2 = ... = \mu_k$ or equivalently, H_o: $\mu_k = 0$ for all treatments (factor levels)

H_1: not all $\mu_1 = \mu_2 = ... = \mu_k$ H_1: $\mu_k \neq 0$ for some treatments

When the *F* obtained is known, the *F* critical needs to be found, the *F* statistic is always positive. There is no one – or two – direction tests for the ANOVA, there is only one *F* critical value and it is always positive. To get the *F* critical value you use the *F* table. Once you have the *F* obtained and the *F* critical values, you simply compare the values just like a *t* test if the *F* obtained value is less than the *F* critical value, the groups are all the same; the null hypothesis is supported; the experimental hypothesis or alternative hypothesis is not supported; and the independent variable did not work, this means that nothing happened in the experiment and the statistical analysis is complete. If the *F* obtained is larger than the *F* critical value, the groups are not all the same, the experimental hypothesis is supported; the null hypothesis is not supported and the independent variable worked, this means that something is going on in the experiment at least two groups are different from each other.

The Wilcoxon Matched – Pairs Signed Rank Test

This is a non-parametric test for a single sample or for paired scores. It is the non-parametric analogue of the *t* test for related sample. It is used to test the hypothesis that two paired samples have come from the same population. Because it is non parametric, it makes only limited assumptions about the distribution of the data.

Kazmier (1996) put it this way; that because the Wilcoxon test considers the magnitude of the difference between each sample value and the hypothesized value of the median, it is a more sensitive test than the sign test. On the other hand, because differences are determined, the values must be at least at the interval scale.

The main ideas behind the Wilcoxon test as suggested by Cardinal (2004) is that, the null hypothesis is that both members of each pair are drawn from the same distribution even though the distribution may vary from pair to pair. Therefore if the null hypothesis is true, the differences are equally likely to be positive or negative and should be distributed symmetrically.

The Mann-Whitney U Test (For 2 Independent Samples) /the Wilcoxon Rank Sum Test

This is a non-parametric equivalent of the unpaired t-test. It is used to test the hypothesis that two independent samples have come from the same population. Because it is non parametric, it makes only limited assumptions about the distribution of the data.

Basically the independent groups design has two groups. These are called experimental and control. In this situation subjects are randomly selected from the population and randomly assigned to two groups. We should take note while using this method there is no basis for pairing scores. Nor is it necessary to have the same number of scores in the two groups.

The basic logic of the test is that, suppose we have two samples with n_1 and n_2 observations in each $\left(n_1 + n_2 = N \text{observation in total}\right)$. We can rank them, lowest to highest, from 1 to N. If the two samples come from identical populations, the sum of the ranks of sample 1 scores is likely to be about the same as the sum of the ranks in sample 2 scores. If, on the other hand sample 1 comes from a population with generally much lower values than sample 2, then the sum of the ranks of sample 1 scores will be lower than the sum of the ranks of sample 2 scores.

There are actually two tests based on the logic used for Mann–Whitney U test . They are the Mann – Whitney U test itself and the Wilcoxon rank- sum test. As the two tests are directly equivalent, they

both give the same p value. The Mann –Whitney U test is more popular and has a name that is not so easily confused with the Wilcoxon signed rank test.

Procedure in calculating the Mann –Whitney U statistic.

(1) Call the smaller group 'group1' and the larger group 'group2', so $n_1 < n_2 \left(if\ n_1 = n_2,\ \text{ignore this step}\right)$.

(2) Calculate the sum of the ranks of group 1 (= R_1) and group 2 (=R2)

(3) $U_1 = R_1 - \frac{n_1\left(n_1+1\right)}{2}$

(4) $U_2 = R_2 - \frac{n_2\left(n_2+1\right)}{2}$

(5) The Mann-Whitney statistic U is the smaller of U_1 and U_2

To be sure of our calculations we can verify the following $U_1 + U_2 = n_1 n_2$ and $R_1 + R_2 = \frac{\left(n_1+n_2\right)\left(n_1+n_2+1\right)}{2}$

It does not matter which numbers you call U_1 and U_2.

If n_2 is small, look up the critical value of U in tables. Values of U smaller than the critical value are significance.

If however, $n_2 > 20$ the U statistic is approximately normally distributed,

Mean $\mu = \frac{n_1 n_2}{2}$

Variance $\sigma^2 = \frac{n_1 n_2\left(n_1+n_2+1\right)}{12}$. So we calculate a Z score.

$$Z = \frac{U - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2\left(n_1+n_2+1\right)}{12}}}$$

We should note that there are three different ways of calculating the Mann-Whitney U test. Here we shall adopt the method discussed above.

Kruskal-Wallis Test

This test is an alternative to the independent group analysis of variance, when the assumption of normality or equality of variance is not met. Like most nonparametric tests, the test assumes that the variable under study has an underlying continuous distribution. It requires at least ordinal measurement of that variable.

The Kruskal-Wallis one-way analysis of variance by ranks is an extremely useful test for deciding whether K independent samples are from different population (Siegel, 1956).

The Kruskal Wallis technique tests the null hypothesis that the K samples come from the same population or from identical population with respect to average.
Siegel (1956, pp186) explained that, in the computation of the Kruskal-Wallis test, each of the observations are replaced by ranks. That is, all of the scores from all of the K samples combined are ranked in a single series. The smallest score is replaced by rank 1, the next smallest by rank 2, and the largest by rank N. N= the total number of independent observations in the K samples.
When this has been done, the sum of the ranks in each sample (column) is found. The Kruskal-Wallis test determines whether these sums of ranks are so disparate that they are not likely to have come from samples which were all drawn from the same population.
We note here, that if the null hypothesis is true, the average of the ranks for each sample groups should be about equal. The test statistic calculated is designated H and is based on the sum of the ranks in each of the several random samples.

$$H = \frac{12}{N(N+1)}\left(\sum_{j=1}^{k}\frac{R_j^2}{nj}\right) - 3(N+1)$$

Where

N = Combined sample size of the several samples
R_j = Sum of the ranks for the j^{th} sample
n_j = Number of observations in the j^{th} sample
K = Number of sample groups

Given that the size of each sample group is at least $n_j \geq 5$ and the null hypothesis is true, Kazmier (1996) and Siegel (1956) believed that the sampling distribution of H is approximately distributed as chi-square with degrees of freedom = k-1 where k is the number of sample groups. The chi-square value that approximates the critical value of the test statistic is always the upper tail value. This test procedure is analogous to the upper tail of the F distribution being used in the analysis of variance.
Kazmier (1996) gave a working formula when there are tied ranks. The test statistic should be corrected and the corrected value of the test statistic is design

$$H_c = \frac{H}{1 - \left[\sum \frac{\left(t_j^3 - t_j\right)}{\left(N^3 - N\right)}\right]}$$

Where t_j represents the number of tied scores in the j^{th} sample

The effect of this correction is to increase the value of the calculated H statistic. Wea Hajek (1964) went ahead to observe that rejection of H_o tells you that at least one of the K samples is drawn from a population with a median different from the others. But it does not tell you which one, or how many are different.

IJMASR: E-ISSN 3027-2122 P-ISSN 3027-0545

Another important suggestion on the use of Kruskal-Wallis one way analysis of variance is that it can be used with K independent groups, where $k \geq 3$, but however, when $k = 2$, you would use the Mann-Whitney–U test instead Hajek, (1964). Also take note that because the samples are independent, they can be of different sizes.

Materials Method

The purpose of any research can be categorized into three groups, Motulsky (2014). It may be to compare two independent samples or examine a set of differences. Another may be comparing three or more groups. And the third may be to assess the linear association between two variables. From these three groups one from each is considered. Those considered are:

i) the unpaired t-test, to tests the hypothesis that the mean examination scores of ND II Statistics students is equal to the mean examination scores of ND I Statistics .The non-parametric equivalent is the Mann-Whitney U-test.

Two Sample Data

In this study, examination scores for the 2005/2006 academic session of the Ramat Polytechnic Maiduguri, Borno State were considered. The examination results consist of 40% as continuous assessment and 60% as examination score. The performance of National Diploma II Statistics students was compared with that of National Diploma I Statistics students.

The aim of this experiment is to test hypothesis for difference between mean scores of the two groups of students.

Case of Large Samples

A random sample of 35 out of the 86 students of ND II Statistics was selected and another random sample of 33 out of 67 students of ND I Statistics was also selected. Random numbers were used in the selection. The performance of each set of students in Business statistics was recorded.

Table 1. The following are the samples taken for both groups:

Sample A (ND II Statistics)					**Sample B (ND1 Statistics)**				
38,	**43,**	**63,**	**64,**	**45**	**54,**	**34,**	**43,**	**65,**	**30**
41,	**55,**	**56,**	**47,**	**51**	**50,**	**40,**	**70,**	**49,**	**28**
40,	**49,**	**67,**	**51,**	**57**	**48,**	**76,**	**40,**	**44,**	**56**
70,	**52,**	**49,**	**49,**	**47**	**45,**	**50,**	**53,**	**57,**	**48**
44,	**46,**	**47,**	**57,**	**37**	**60,**	**63,**	**61,**	**51,**	**40**
45,	**37,**	**66,**	**18,**	**51**	**43,**	**49,**	**41,**	**58,**	**70**
49,	**51,**	**65,**	**52,**	**55**	**64,**	**58,**	**45**		

Case of Small Samples

A random sample of 17 out of 86 ND II Statistics students and another random sample of 15 out of 67 students of ND I Statistics were selected. The selection was by using random numbers

IJMASR: E-ISSN 3027-2122 P-ISSN 3027-0545

Table 2. The following are the samples taken for both groups:

Sample A (ND II Statistics)					**Sample B (ND I Statistics)**				
57,	**68,**	**75,**	**41,**	**47**	**61,**	**63,**	**48,**	**38,**	**40**
22,	**49,**	**67,**	**37,**	**55**	**34,**	**61,**	**51,**	**42,**	**76**
63,	**65,**	**60,**	**18,**	**44**	**64,**	**53,**	**51,**	**50,**	**40**
73,	**53**								

RESULTS AND DISCUSSION

Results

Case of Large Samples (Parametric Analysis)

The application of any parametric test requires certain parametric assumptions. However for a large sample i.e. $n > 30$, the central limit theorem stated that the data set approximates to the standard normal variable. Here the appropriate parametric test to use for testing the difference of means will be the *t*-test. Thus without making any assumption the *Z* statistic gives the same result as the *t*-statistic.

The Hypothesis test $H_0 : \mu_A = \mu_B$ against $H_1 : \mu_A \neq \mu_B$

The output of two sample t-test using R is given below.

Variable	N	Mean	StDev	SE Mean	95.0% CI
x	35	50.11	10.29	1.74	(46.58, 53.65)
y	33	51.00	11.47	2.00	(46.93, 55.07)

Two-sample t test power calculation

n	=	35
delta	=	0.89
sd	=	1
sig.level	=	0.05
power	=	0.9563757
alternative	=	two.sided

The R package consider the two missing values therefore *n*=35

Case of Large samples (Nonparametric Analysis)

The same data set used above for the parametric analysis is to be used for non-parametric analysis. Here the appropriate corresponding non-parametric test is the Mann-Whitney U test. The non-parametric test does not require any parametric assumption.

The Hypothesis test

$H_0 : \mu_A = \mu_B \ against \ H_1 : \mu_A \neq \mu_B$

The output of the Mann Whitney U test using R is given below.

GrpA N = 33 Median = 50.000

GrpB N = 35 Median = 49.000

Point estimate for ETA1-ETA2 is -0.000

95.0 Percent CI for ETA1-ETA2 is (-5.004, 6.000)

W = 1151.5
Test of ETA1 = ETA2 vs. ETA1 not = ETA2 is significant at 0.8781
The test is significant at 0.8780 (adjusted for ties)
Cannot reject at alpha = 0.05
The power for adjusted ties is 0

Case of Small Samples (Parametric Analysis)

The appropriate parametric test to apply is the unpaired *t* test. Note that the application of any parametric test require certain parametric assumptions therefore for the fulfillment of the normality assumption, we construct Histogram of the two groups.

Hypothesis test

$$H_0 : \mu_A = \mu_B \ against \ H_1 : \mu_A \neq \mu_B$$

The output of the sample t-test using R is given below.

Data: x by y
t = 0.2238, df = 28.821, p-value = 0.8245
Alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-9.132566 11.375703
Sample estimates:
Mean in group 1 mean in group 2
52.58824 51.46667
Two-sample t test power calculation
n = 17
Delta = 1.12
sd = 1
sig.level = 0.05
Power = 0.8860768
Alternative = two.sided

The R package consider the missing values therefore n=17

Case of Small Samples (Nonparametric Analysis)

The same data set used above for the parametric analysis (small samples) is to be used for non-parametric analysis.

Here the appropriate corresponding non-parametric test is the Mann-Whitney U test.

$$H_0 : \mu_A = \mu_B \ against \ H_1 : \mu_A \neq \mu_B$$

The output of the Mann Whitney U tests using R given below.

smpA N = 15 Median = 51.00
smpB N = 17 Median = 55.00
Point estimate for ETA1-ETA2 is -3.00
95.0 Percent CI for ETA1-ETA2 is (-13.00,8.01)
W = 232.0

Test of ETA1 = ETA2 vs ETA1 not = ETA2 is significant at 0.5711
The test is significant at 0.5709 (adjusted for ties)
Cannot reject at alpha = 0.05
The power for adjusted ties is 0.5709

SUMMARY AND CONCLUSION

Considering large samples, the parametric *t*-test generated a power of 95.64%, while the non-parametric Mann-Whitney *U*-test recorded a power of 87.80%. Also looking at the small samples, the parametric *t*-test has a power of 86.61% and the Mann-Whitney *U*-test with a power of 57.01%.

The one-way ANOVA has the same power of 85.06% as the non-parametric Kruskal-Wallis when large samples are considered. Also on small samples, the two tests generated the same amount of power of 30.92%.

In the case of correlation analysis, the parametric Pearson product moment correlation gave a power of 61.01% and the Spearman rank correlation recorded a power of 60.36%. Considering small samples, the parametric tests generated a power of 37.76% and the non-parametric counterpart gave a power of 19.86%.

From the above calculations of statistical power, we observe as follows.

i) When large samples are considered parametric tests has an average power of 80.57% while the non-parametric counterpart gave an average power of 77.74%. This shows that parametric tests has more power of about 2.83% when large samples are considered.

ii) When small samples are considered, parametric tests gave an average power of 52.43% while the non-parametric counterparts recorded an average power of 35.76% which indicate that parametric tests has more power of about 16.67% over the non-parametric tests when small samples are considered.

iii) Considering both large and small samples combined, the parametric tests gave an average power of 66.5% while the non-parametric counterparts recorded an average power of 56.85%. This shows that on the overall, the parametric test has more power of 9.65% over the non-parametric tests.

REFERENCES

Bewick, V., Cheek, L., & Ball, J. (2004). Statistics review 10: further nonparametric methods. *Critical care*, *8*, 1-4.

Cann, C. G., Shen, C., LaPelusa, M. B., Agarwal, R., Cardin, D. B., & Eng, C. (2022). Supportive care (SC) utilization for patients with locally advanced pancreatic cancer: Review of the National Cancer Data Base (2004-2018). *Journal of Clinical Oncology*, *40*(16_suppl), 4154-4154.

Dalmaijer, E. S., Nord, C. L., & Astle, D. E. (2022). Statistical power for cluster analysis. *BMC bioinformatics*, *23*(1), 205.

Davies, W. T., Myer, G. D., & Read, P. J. (2020). Is it time we better understood the tests we are using for return to sport decision making following ACL reconstruction? A critical review of the hop tests. *Sports medicine*, *50*, 485-495.

Di Leo, G., & Sardanelli, F. (2020). Statistical significance: p value, 0.05 threshold, and applications to radiomics—reasons for a conservative approach. *European radiology experimental*, *4*, 1-8.

Frank, H. & Altheon, S.C. (1995), *Statistics. Concepts and Application*. Low Price Edition, Cambridge University Press.

Freund, J.E. and Walpole, R.E. (1980), *Mathematical Statistics*, 3rd Edition Prentice Hall International, Inc. London.

Fruend, J. E. & Williams, F. J. (1988). *Elementary Business Statistics*: *The Modern Approach* 4th Edition. Prentice-Hall International Inc.

Hajek, J. (1964). Asymptotic theory of rejective sampling with varying probabilities from a finite population. *The Annals of Mathematical Statistics*, *35*(4), 1491-1523.

Havre, S., Hetzler, E., Whitney, P., & Nowell, L. (2002). Themeriver: Visualizing thematic changes in large document collections. *IEEE transactions on visualization and computer graphics*, *8*(1), 9-20.

Huntsberger, D.V., Billingsley, P., (1981). Elements of Statistical Inference, fifth ed. Allyn Bacon, Inc., Boston.

Joshi, R. D., & Dhakal, C. K. (2021). Predicting type 2 diabetes using logistic regression and machine learning approaches. *International journal of environmental research and public health*, *18*(14), 7346.

Karunasingha, D. S. K. (2022). Root mean square error or mean absolute error? Use their ratio as well. *Information Sciences*, *585*, 609-629.

Kazmier, L. J. (1996). Schaum's Outline of Business Statistics 3 rd Edition Mc Graw-Hill New York.

Kelton, D.W. (2000), *Experimental Design for Simulation*. Proceedings of the 2000 Winter Simulation Conference. Department of Qualitative Analysis and Operation Management, College of Business Administration, University of Cincinnati; OH 45221-0130, USA.

Kilty, K. (2001), *Design Your Experiments, http://www.sas.org/E-Bulletin/2001-12-14/toc.html*.

Kim, J. S. & Kalb, J. W. (2002). Design of Experiments: *An Overview and Application Example*. Medical Device Diagnosis Industry Magazine MDDI article CANON Communication LLC 2002.

Lan, Y., Lian, Z., Zeng, R., Zhu, D., Xia, Y., Liu, M., & Zhang, P. (2020). A statistical model of the impact of online rumors on the information quantity of online public opinion. *Physica A: Statistical Mechanics and its Applications*, *541*, 123623.

Lee-Yaw, J., L. McCune, J., Pironon, S., & N. Sheth, S. (2022). Species distribution models rarely predict the biology of real populations. *Ecography*, *2022*(6), e05877.

Lovrich M. (2002) Article on Non-parametric Methods Week 12b. University of South Australia.

Luengo, D., Martino, L., Bugallo, M., Elvira, V., & Särkkä, S. (2020). A survey of Monte Carlo methods for parameter estimation. *EURASIP Journal on Advances in Signal Processing*, *2020*, 1-62.

Maier, M., & Lakens, D. (2022). Justify your alpha: A primer on two practical approaches. *Advances in Methods and Practices in Psychological Science*, *5*(2), 25152459221080396.

Maxwell, S. E., & Delaney, H. D. (1990). *Designing Experiments and Data Analysis: a Model Comparison Perspective*. Wadsworth.

Motulsky, H. (2014). *Intuitive biostatistics: a nonmathematical guide to statistical thinking*. Oxford University Press, USA.

Noble, S., Scheinost, D., & Constable, R. T. (2021). Cluster failure or power failure? Evaluating sensitivity in cluster-level inference. *Neuroimage*, *209*, 116468.

Pattern, M.L. (2002), *Understanding Research Methods: An Overview of the Essentials,* (3rd Edition), Los Angeles: Pyrc Zak Publishing.

Scheel, A. M., Tiokhin, L., Isager, P. M., & Lakens, D. (2021). Why hypothesis testers should spend less time testing hypotheses. *Perspectives on Psychological Science*, *16*(4), 744-755.

Serdar, C. C., Cihan, M., Yücel, D., & Serdar, M. A. (2021). Sample size, power and effect size revisited: simplified and practical approaches in pre-clinical, clinical and laboratory studies. *Biochemia medica*, *31*(1), 27-53.

Sheskin, D. J. (2003). *Handbook of parametric and nonparametric statistical procedures*. Chapman and hall/CRC.

Siegel S. and Castellan N.J. (1988) Non-parametric Statistics for the Behavioural Science 2nd Edition New York: McGraw. Hill.

Smith, R. J. (2020). P>. 005: The incorrect interpretation of "not significant" results is a significant problem. *American journal of physical anthropology*, *172*(4), 521-527.

Stapor, K., & Stapor, K. (2020). Descriptive and inferential statistics. *Introduction to Probabilistic and Statistical Methods with Examples in R*, 63-131.

Tabachnick, B. G., & Fidell, L. S. (2007). *Experimental designs using ANOVA* (Vol. 724). Belmont, CA: Thomson/Brooks/Cole.

Terrell, S. R. (2021). *Statistics translated: A step-by-step guide to analyzing and interpreting data*. Guilford Publications.

The New Encyclopedia Britannica (1998). Statistics Macropaedia Knowledge in-depth Vol. 28, 15th Edition, by encyclopedia Britannica Inc.

Tintle, N., Chance, B. L., Cobb, G. W., Rossman, A. J., Roy, S., Swanson, T., & VanderStoep, J. (2020). *Introduction to statistical investigations*. John Wiley & Sons.

Townsend, J., Courchesne, E., Covington, J., Westerfield, M., Harris, N. S., Lyden, P., & Press, G. A. (1999). Spatial attention deficits in patients with acquired or developmental cerebellar abnormality. *Journal of Neuroscience*, *19*(13), 5632-5643.

Trafimow, D., Hyman, M. R., Kostyk, A., Wang, C., & Wang, T. (2021). The harmful effect of null hypothesis significance testing on marketing research: An example. *Journal of Business Research*, *125*, 39-44.

Tredennick, A. T., Hooker, G., Ellner, S. P., & Adler, P. B. (2021). A practical guide to selecting models for exploration, inference, and prediction in ecology. *Ecology*, *102*(6), e03336.

Tredennick, A. T., Hooker, G., Ellner, S. P., & Adler, P. B. (2021). A practical guide to selecting models for exploration, inference, and prediction in ecology. *Ecology*, *102*(6), e03336.

Wasserman, L., Ramdas, A., & Balakrishnan, S. (2020). Universal inference. *Proceedings of the National Academy of Sciences*, *117*(29), 16880-16890.

Yao, L., Chu, Z., Li, S., Li, Y., Gao, J., & Zhang, A. (2021). A survey on causal inference. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, *15*(5), 1-46.