THE INTERNATIONAL JOURNAL OF SCIENCE & TECHNOLEDGE

Relative Efficiency of Some Tests of Two Population Means under Violation of Homoscedacity Assumptions

Paul Moses Medugu
Lecturer, Department of Mathematics/Statistics,
Federal Polytechnic Mubi, Adamawa State, Nigeria
Yusuf Abubakar Mohammed
Lecturer, Department of Mathematical Science,
University of Maiduguri, Borno State, Nigeria
Haruna Bakari Rann
Lecturer, Department of Mathematical Science,
University of Maiduguri, Borno State, Nigeria

Abstract:
Researchers have been using statistical tools for data analysis that are based on the assumption of normality and other assumptions related to specific tools. When the researchers have developed the theoretical aspects related to these methods, they have assumed that the population from where the sample has been drawn follows a normal law. But, in reality the population need not be always normal or homoscedastic. When these methods are used for non-normal population, they may lead to unreliable results and also inferences with low power. This study therefore investigates two parametric tests (t and Welch's t tests) and their performances are compared with nonparametric counterparts of testing two independent population of unequal variances from normal, uniform, gamma and exponential distributions at different sample sizes. The results of the analysis revealed that the Welch's t- test is the best on data from normal distribution, Mann-Whitney from uniform, and Median from gamma and exponential distributions.

Keywords: *t-test, Welch tests, median test, man-Whitney U test*

1. Introduction

A common problem in applied research is to decide whether or not sample differences in central tendency reflect true differences in parent populations. It is appropriate to use t-test for two sample case (two groups) if assumptions of normality, homogeneity of variance and independent of errors are met. However, when these assumptions are violated, the nonparametric equivalent is considered (Akeyede et al, 2014 and Keselman et al, 2004). The issue of handling uncertainties can be a major issue. For parametric tests, some assumptions need to be satisfied like normality, sample size etc. Whereas, non-parametric tests does not strictly requires any kind of assumption testing. Non-parametric methods that readily suggest themselves include the Median and the Mann-Whitney U test (Gibbon, 1992; Afuecheta et al, 2012). Statistical tests could be for either paired or unpaired wherein each subject has two measurements which are done before and after the treatment (Rasch et al, 2010 and Rochon 2012). The methods/techniques are: t-test and welch t-test from parametric procedure, and Median test and Mann-Whitney U test from non-parametric procedure. All the above listed methods are for test of significant difference between variables/subjects when having independent samples (Edith and Nkiru, 2016). In statistical computation, test statistics sometimes are affected by nature of data (Derrick et al, 2016). In this work, methods of analyzing two independent samples drawn from independent populations would be considered by subjecting some set of data to different conditions, such as sample size, to determine the condition under which they perform optimally in terms of Relative Efficiency (R.E) and power. (Mood, 1954 and Srilakshminarayana, 2015)
The study is therefore aimed at investigating the efficiency and relative performance of four test statistics to ascertain which test is most efficient under certain conditions and distributions.

2. Methodology

In this study, data were generated using simulation procedures from Normal, Uniform, Exponential and Gamma distributions. The research work therefore, considered the parametric (Student t-test and Welch's t-test) and nonparametric (Median test and Mann-Whitney U test.

3. Simulation Procedures and Analysis

Random samples were simulated from Normal, Uniform, Gamma and Exponential distributions respectively for sample size of 5, 10, 15, 20, 25, 30, 35, 40, 45 and 50. Each test procedures were applied on the data sets at varying sample sizes. The processes are repeated 3000 times.

3.1. Student t-test

Independent sample t-test: here, we consider two situations to test the difference between two population means, when the samples have been drawn independently. In the first situation, we will consider the datasets that are generated from a normal distribution and in the second situation, datasets that are generated from a non-normal distribution. Then the statistical null hypothesis isH_0 :μ_1 = $\mu_2$that this hypothesis is true, if and only if μ_1 - μ_2 = 0, so in a sense it is a hypothesis about the mean difference. We are interested in whether or not μ_1 - μ_2 is equal to zero. So, we take independent samples of size n_1 and n_2, and compute sample means $\bar{x}_1$and $\bar{x}_2$. We then examine$\bar{x}_1$ - $\bar{x}_2$ and see if it is "different enough from zero to be statistically significant

Student's *t*-test assumes that the two populations have normal distributions with equal variances. The data are continuous and are measured with interval scale. The numerator includes two sample statistics, but ultimately reduces them to a single number, $\bar{x}_1$ - $\bar{x}_2$. We standardize the mean difference $\bar{x}_1$ - $\bar{x}_2$ by subtracting its null-hypothesized mean, and dividing by its standard error. It turns out that the standard error of $\bar{x}_1$ - $\bar{x}_2$ is given by equation 1, if we assumed equal variances, one way of approaching this is to simply drop the subscript on $\sigma_1^2$and σ_2^2, since they are now assumed to be the same σ^2.

$$SE(\bar{x}_1 - \bar{x}_2) = \sqrt{\sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} \qquad 1$$

To test whether there is a difference or not between the means of the two populations

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim t_{n_1+n_2-2} \qquad 2$$

There are many possible estimates of σ^2 that we could construct from the two sample variances. Remember, we are assuming each of the two sample variances is estimating the same quantity. It turns out that one particular estimate, if substituted for σ^2, yields a statistic that has an exact Student t distribution.

This estimate goes by a variety of names. It is sometimes called "the pooled unbiased estimator," and also sometimes called "Mean Square Within" or "Mean Square Error. The specific formula for two groups is;

$$\hat{\sigma}^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \qquad 3$$

Note that the resulting t-statistic has n_1 +n_2−2 degrees of freedom

3.2. Welch's t-test

In statistics, Welch's *t*-test is a two-sample location test which is used to test the hypothesis that two populations have equal means. It is named for its creator, Bernard Lewis Welch, and is an adaptation of Student's t-test, and is more reliable when the two samples have unequal variances and/or unequal sample sizes. These tests are often referred to as "unpaired" or "independent samples" *t*-tests, as they are typically applied when the statistical units underlying the two samples being compared are non-overlapping. Given that Welch's *t*-test has been less popular than Student's *t*-test and may be less familiar to readers, a more informative name is "Welch's unequal variances *t*-test" or "unequal variances *t*-test" for brevity

Student's *t*-test assumes that the two populations have normal distributions with equal variances. Welch's *t*-test is designed for unequal variances, but the assumption of normality, continuous data and interval scales are maintained. Welch's *t*-test is an approximate solution to the

$$teststatistic = \frac{estimate - hypothesisedvalue}{standarderror} \qquad 4$$

In the case of two independent samples, our estimate is the difference between the two-sample means. The hypothesized value is the difference we hypothesize between the two true population means this is often zero (to test whether there is a difference or not between the means of the two populations).

$$t^{'} = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{SE} \qquad 5$$

In a Welch's two independent sample t-test, we estimate our standard error using the sample standard deviations.

$$SE = \sqrt{\frac{s_1^2}{n_1} + \frac{s_1^2}{n_1}} \qquad 6$$

Where, $\bar{x}_1 and \bar{x}_2$ are the 1st and 2nd sample means, $s_1^2 and s_2^2$ are sample variances and $n_1 and n_2$ sample sizes. Unlike in Student's *t*-test, the denominator is *not* based on a pooled variance estimate.

The degrees of freedom associated with this variances estimates is approximated using the Welch–Satterthwaite equation The degree of freedom of the statistic is obtained as follows

$$df^{'} = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_1^2}{n_1}\right)^2}{\frac{s_1^4}{n_1^2(n_1-1)} + \frac{s_2^4}{n_2^2(n_2-1)}} \qquad 7$$

Statistical packages estimate the degrees of freedom for the Welch's t-test, or more simplistically (and less accurately) we can estimate the degrees of freedom by subtracting one from the smaller of the two sample sizes. Once *t* and $t^{'}$ have been computed, these statistics can be used with the *t*-distribution to test one of two possible null hypotheses:

that the two-population means are equal, in which a two-tailed test is applied; or
that one of the population means is greater than or equal to the other, in which a one-tailed test is applied.

3.3. Median Test

The median test is a statistical procedure for testing whether two independent populations differ in their measure of central tendency or location (Oyeka et al. 2009 and 2010).

- Observations are independent
- Observations come from populations with continuous distribution functions

Let x_i be the i observation in a random sample of size $n_j$that the two populations are measured on at least the ordinal scale. To apply the two-sample median test by ranks, we first pool the two samples into one combined sample of size $n = \sum_{j=1}^{2} nj; j = 1,2 = n_1 + n_2$ observations in the pooled sample are now ranked, either from the largest to the smallest or from the smallest to the largest. Now under the hypothesis of equal population medians, then in the absence of ties, any one randomly selected observation in the combined sample is as likely to be greater as less than any other observation in the sample, and hence, is equally likely to receive any one of the ranks assigned to the observations, thereby justifying the use of the median ranks test for two populations. Let $R_j = \sum_{i=1}^{n_j} r_{ij}$ 8

Be the sum of the ranks assigned to observations drawn from population j for j=1, 2, with mean rank $\overline{r_{.j}} = \frac{R_j}{n_j} = \sum_{i=1}^{n_j} \frac{r_{ij}}{n_j}$ 9

The overall mean rank is, $\bar{r} = \frac{n(n+1)}{2n} = \frac{n+1}{2}$ 10

The total variance of all the ranks is, $\sigma^2 == \frac{n(n+1)}{12}$ 11

Now the sum of squared deviations of observed sample or treatment group mean rank $\overline{r_{.j}}$ from their overall mean rank $\bar{r}$ is $S_{ab}^2 = \sum_{j=1}^{2} n_j (\overline{r_{.j}} - \bar{r}) = \sum_{j=1}^{2} \frac{R_{.j}^2}{n_j} - \frac{n(n+1)^2}{4}$ now the quadratic form,

$$Q = \chi^2 = \frac{S_{ab}^2}{\sigma^2} = \frac{\sum_{j=1}^{2} \frac{R_{.j}^2}{n_j} - \frac{n(n+1)}{4}}{\frac{n(n+1)}{12}} \qquad 12$$

That is,

$$Q = \chi^2 = \frac{12}{n(n+n1)} \sum_{j=1}^{2} \frac{R_{.j}^2}{n_j} - 3(n+1) \qquad 13$$

has approximately a chi-square distribution with k-1=2-1=1 degree of freedom for sufficiently large n [3,4], and may be used to test the null hypothesis of equal population medians. The null hypothesis is rejected at theα level of significance if $Q = \chi^2 \geq \chi_{1-\alpha,1}^2$ otherwise the null hypothesis is accepted. Note that equation 5 can be alternatively expressed as

$$S_{ab}^2 = \frac{(n_1 R_2 - n_2 R_1)^2}{n_1 n_2 (n_1 + n_2)^2} \qquad 14$$

Hence, the test statistic of equation 6 can be equivalently written as:

$$Q = \chi^2 = \frac{12((n_1 R_2 - n_2 R_1)^2)}{n_1 n_2 (n_1 + n_2)^2 (n_1 + n_2 + 1)} \qquad 15$$

The test statistic of equation 6 and 8 are sufficiently adequate and yields good results, provided $n_1$and $n_2$are each at least 5 [2]. Now if the two sample are equal so that $n_1=n_2=m$ says, then equation 8 further reduces to

$$Q = \chi^2 = \frac{12((n_1 R_2 - n_2 R_1)^2)}{m(m)^2 (m+1)} \qquad 16$$

3.4. Mann-Whitney U test

Mann-Whitney U test is the non-parametric alternative test to the independent sample t-test. There are, however, some assumptions are as follows:

- All the observations from both groups are independent of each other.
- The responses are ordinal (i.e., one can at least say, if any two observations, which is the greater).
- The distributions of both groups are equal under the null hypothesis, so that the probability of an observation from one population (X) exceeding an observation from the second population (Y) equals the probability of an observation from Y exceeding an observation from X. That is, there is asymmetry between populations with respect to probability of random drawing of a larger observation.

(iv). under the alternative hypothesis, the probability of an observation from one population (X) exceeding an observation from the second population (Y) (after exclusion of ties) is not equal to 0.5. The alternative may also be stated in terms of a one-sided test, for example: P(X > Y) + 0.5 P(X = Y) > 0.5.

$$U = n_1 n_2 - \frac{n_1(n_2+1)}{2} - \sum_{i=r_1+1}^{r_2} R_i \qquad 17$$

Where:

$U = Mann - Whitney U test$

$n1 = sample size one$

$n2 = sample size two$

$R_i = Rank of the sample size$

4. Analysis and Results

Type I error was computed for each test on simulated data from different distributions and sample sizes in such a way that the p-value of rejecting null hypothesis of equal mean used to generate data are recorded. The test with lowest type I error rate is considered as the best at a particular category. In the case of power, different means was used to generate data and power was computed for each test on simulated data from different distributions and sample sizes in such a way that the p-value of rejecting the wrong fixed null hypothesis of equal mean in order to accept the right alternative are counted as power and recorded. The test with highest rejection rate of wrong null hypothesis is considered as the one with highest power and best at a particular category

Sample Size	Test Statistics			
	T-test	Welch's t- test	Median Test	Mann-Whitney U test
5	0.6699	0.6718	0.5000	0.8413
10	0.1618	0.1714	0.9893	0.9655
15	0.0983	0.1002	0.6964	0.8607
20	0.9115	0.9119	0.2517	0.3689
25	0.9977	0.9977	0.5000	0.7730
30	0.7730	0.7733	0.7077	0.8776
35	0.4405	0.4417	0.1553	0.2189
40	0.6803	0.6806	0.8659	0.9885
45	0.6222	0.6226	0.8144	0.6645
50	0.2204	0.2220	0.3359	0.4219

Table 1: Relative Frequency of Type I error for Unequal Variance ($\delta_1^2 \neq \delta_2^2$) from Normal Distribution

Table 1 shows the relative performance of the four tests using type I error as a criterion for the assessment, when the variances of the two-sample data generated from normal are not equal.Welch's t- test is the best as it has the values far from type I error. This is followed by the parametric t- test. The Mann-Whitney test is better than median test.

Sample Size	Test Statistics			
	T-test	Welch's t- test	Median Test	Mann-Whitney U Test
5	0.2287	0.2287	0.9688	0.9795
10	0.9330	0.9331	0.6230	0.9705
15	0.0084	0.0096	0.9824	0.0235
20	0.0164	0.0165	0.9941	0.0283
25	0.1190	0.1193	0.8852	0.1413
30	0.0953	0.0957	0.8192	0.1026
35	0.0478	0.0480	0.9552	0.0649
40	0.0283	0.0286	0.9597	0.0545
45	0.2342	0.2344	0.8837	0.2433
50	0.1090	0.1092	0.8389	0.1566

Table 2: Relative Frequency of Type I error for Unequal Variance ($\delta_1^2 \neq \delta_2^2$) from Uniform Distribution

Table 2 shows that the two non-parametric tests are farer from type I error of α = 0.05 than parametric counterparts at various sample sizes, where the Mann-Whitney test is the best among them at smaller sample sizes (5 and 10) while the median test is the best as sample size is getting larger. The t-test has the closest values of type I error at various sample sizes and hence is the worst among the four tests followed by the Welch parametric test.

Sample Size	Test Statistics			
	T-test	Welch's t- test	Median test	Mann-Whitney U test
5	0.2812	0.3121	0.9688	0.1508
10	0.3189	0.3321	0.9990	0.0115
15	0.3403	0.3484	0.9408	0.0367
20	0.3089	0.3153	0.9423	0.3408
25	0.1502	0.1567	0.9980	0.0247
30	0.3857	0.3879	0.9919	0.0284
35	0.3183	0.3219	0.9795	0.1631
40	0.5523	0.5533	0.9989	9.707e-05
45	0.5206	0.5221	1.0000	3.511e-07
50	0.2040	0.2069	0.9923	0.0001

Table 3: Relative Frequency of Type I error for Unequal Variance ($\delta_1^2 \neq \delta_2^2$) from Gamma Distribution

From table 3, the median test is the best among the four tests at all sample sizes for data simulated from Gamma distribution where variances are not equal followed by Welch's t-test. The least performing test in this case is the Mann-Whitney U test due to its values very close to type I error. It also observed that Mann Whitney committed full type I error, at 5% level of significance, as sample sizes getting large especially from 30 upward.

Sample Size	Test Statistics			
	T-test	Welch's t- test	Median test	Mann-Whitney U test
5	0.3155	0.3334	0.9688	0.6905
10	0.5122	0.5187	0.7190	0.593
15	0.0116	0.0158	0.9963	0.0178
20	0.1414	0.1489	0.9423	0.7180
25	0.0003	0.0006	0.9927	0.0018
30	1.924e-05	5.491e-05	1	9.037e-07
35	0.0020	0.0026	0.9795	0.0014
40	5.34e-06	1.375e-05	0.9968	6.537e-05
45	7.674e-06	1.555e-05	0.9988	9.909e-06
50	2.821e-06	4.66e-06	1	1.519e-05

Table 4: Relative Frequency of Type I error for Unequal Variance ($\delta_1^2 \neq \delta_2^2$) from Exponential Distribution

It was observed from table 4; that both parametric tests have values farer from the type I error compared with the non-parametric counterpart from sample size of 5 to 25 which categorized them as the best. However, from sample size of 30 to 50, the median test only has the farthest values from type I error and it is the best at that category. Meanwhile, all the tests aside from median test largely committed full type I error as sample size is getting large.

Sample Size	Test Statistics			
	T-test	Welch's t- test	Median Test	Mann-Whitney U test
5	1.046e-07	1.008e-07	1	0.007937
10	3.356e-11	6.262e-12	1	1.083e-05
15	3.212e-16	3.089e-16	1	1.289e-08
20	2.432e-16	2.387e-15	1	1.451e-11
25	<2.2e-16	<2.2e-16	1	1.582e-14
30	<2.2e-16	<2.2e-16	1	<2.2e-16
35	<2.2e-16	<2.2e-16	1	<2.2e-16
40	<2.2e-16	<2.2e-16	1	<2.2e-16
45	<2.2e-16	<2.2e-16	1	<2.2e-16
50	<2.2e-16	<2.2e-16	1	<2.2e-16

Table 5: Relative Frequency of Power of Test for Unequal Variance ($\delta_1^2 \neq \delta_2^2$) from Normal Distribution

Table 5 shows the relative performance of the four tests, using power as a criterion for the assessment, when the variances of the two-sample data generated from normal are not equal. It was observed that that the Welch's t- test has the strongest rejection value to the wrong null hypothesis than other tests and indeed has the highest power. This is followed by the parametric test. However, as the sample size increases, their power becomes stronger and the t-, Welch's and Mann-Whitney U tests have similar power values.

Sample Size	Test Statistics			
	T-test	Welch's t- test	Median Test	Mann-Whitney U Test
5	0.9179	0.9184	0.8125	0.6905
10	0.8551	0.8539	0.6230	0.6030
15	0.8654	0.8646	0.5000	0.4748
20	0.8762	0.8671	0.5881	0.4211
25	0.7699	0.7006	0.7878	0.7437
30	0.4151	0.4083	0.8998	0.1381
35	0.1339	0.1380	0.0204	0.0129
40	0.9669	0.9670	0.3179	0.0886
45	0.7794	0.7800	0.5000	0.4552
50	0.5046	0.5061	0.4439	0.4227

Table 6: Relative frequency of Power of Test for Unequal Variance ($\delta_1^2 \neq \delta_2^2$) from Uniform Distribution

Table 6 shows that the two non-parametric tests have a stronger rejection to wrong null hypothesis and power at $\alpha = 0.05$ than parametric counterparts at various sample sizes, where the Mann-Whitney test is the best among them. As sample size is getting larger, the t-test has the weak power at various sample sizes and hence is the worst among the four tests followed by the Welch parametric test

Sample Size	Test Statistics			
	T-test	Welch's t- test	Median Test	Mann-Whitney U test
5	0.3116	0.3409	0.0009	0.0555
10	0.5254	0.5262	0.0005	0.0010
15	0.1983	0.2087	0.0002	0.0002
20	0.3307	0.3368	2.241e-08	9.249e-07
25	0.5343	0.5346	1.75e-12	1.75e-07
30	0.0610	0.0658	<2.2e-16	4.126e-10
35	0.5100	0.5121	<2.2e-16	4.503e-08
40	0.0278	0.0307	<2.2e-16	9.769e-11
45	0.0021	0.0028	<2.2e-16	<2.2e-16
50	0.3828	0.3848	<2.2e-16	9.878e-10

Table 7: Relative frequency of Power of Test for Unequal Variance ($\delta_1^2 \neq \delta_2^2$) from Gamma Distribution

From table 7, the median test is the best among the four tests at all sample sizes for data simulated from Gamma distribution where variances are not equal followed by Man-Whitney U test. The least performing test in this case is the Welch test due to its weak rejection values. It also observed that Mann Whitney and median tests have the strongest wrong hypotheses, at 5% level of significance, as sample sizes getting large especially from 30 upward, hence have they the highest power at that category.

Sample Size	Test Statistics			
	T-test	Welch's t- test	Median test	Mann-Whitney U test
5	0.2875	0.3187	0.7865	0.8773
10	0.0030	0.0021	0.0007	0.0008
15	0.0001	0.0002	0.0005	3.673e-06
20	0.0001	0.0004	0.0059	3.4523e-05
25	3.120e-06	2.744e-05	2.432e-08	2.112e-12
30	2.781e-06	2.064e-05	2.768e-12	1.307e-11
35	1.987e-07	2.001e-06	<2.2e-16	3.210e-13
40	1.098e-09	1.498e-08	<2.2e-16	4.329e-14
45	1.145e-09	1.241e-07	<2.2e-16	2.452e-13
50	1.044e-09	1.377e-08	<2.2e-16	3.145e-13

Table 8: Relative frequency of Power of Test for Unequal Variance ($\delta_1^2 \neq \delta_2^2$) from Exponential Distribution

It was observed from table 8; that both non-parametric tests have stronger rejection values to wrong null hypothesis compared with the parametric counterpart from sample size of 5 to 50 which categorized them as the best with median test as the best. The performances of all tests increase as sample size increases.

5. Conclusion

It was observed that the Welch's t- test is the best on data from normal distribution followed by the parametric t-test. The Mann-Whitney test is better than median test. The median test is the best at all sample sizes for data simulated from Gamma distribution followed by Welch's t-test. The least performing test in this case is the Mann-Whitney U test due to its values very close to type I error. It also observed that Mann Whitney committed full type I error, at 5% level of significance, as sample sizes getting large especially from 30 upward. It was observed from exponential distribution; that both parametric tests are better than the non-parametric counterparts from sample size of 5 to 25 which categorized them as the best. However, from sample size of 30 to 50, the median test only has two-sample the best performance based on the two criteria of the assessment. Meanwhile, all the tests aside from median test largely committed full type I error as sample size is getting large.

6. References

i. Afuecheta, E.O., Oyeka, C.A., Ebuh, G.U. and Nnanatu, C.C. (2012) Modified Median Test Intrinsically Adjusted for Ties. Journal of Basic Physical Research, 3, 30-34.

ii. Akeyede, I., Usman, M. and Chiawa, M. A. (2014). On Consistency and Limitation of paired t-test, Sign and Wilcoxon Sign Rank Test. IOSR Journal of Mathematics (IOSR-JM).Volume 10, Issue 1 Ver. IV. PP 01-06

iii. Derrick, B., Toher, D., & White, P. (2016). Why Welch's test is Type I error robust. The Quantitative Methods for Psychology, 12(1), 30-38.

iv. Edith U. U. and Nkiru O. E. (2016). Comparison of Two Sample Tests Using Both Relative Efficiency and Power of Test. Open Journal of Statistics, 6, 331-345

v. Gibbon, J.D. (1992). Nonparametric Statistics: An Introduction. Quantitative Applications in Social Sciences, Sage Publications, New York.

vi. Keselman, H., J., Othman, A. R., Wilcox R. R., and Fradette, K. (2004) The New and ImprovedTwo Sample t-Test. Psychological Science 15(1), pp. 47-51.

vii. Mood, A.M. (1954). On the Asymptotic Efficiency of Certain Nonparametric Two-Sample Tests. Annals of Mathematical Statistics, 25, 514-522

viii. Oyeka C.A. (2009). An Introduction to Applied Statistical Methods (4thEdn) Nobem Avocation

ix. Publishing Company, Enugu, Nigeria.

x. OyekaC.A., Utazi CE, Nwosu CR, Ebuh GU, Ikpegbu PA, et al. (2010) A statistical comparisonof Test scores for non-parametric approach. Journal of Mathematics Sciences 21: 77-87

xi. Rasch, D., Kubinger, K. D, &Moder, K. (2011). The two-sample t test: pre-testing its assumptions do not pay off. Statistical papers

xii. Rochon, J., Gondan, M., &Kieser, M. (2012). To test or not to test: Preliminary assessment of normality when comparing two independent samples. BMC Medical Research Methodology, 12(1), 1, Probability Letters, Vol. 18, pp. 241-244

xiii. Srilakshminarayana, G. (2015). On Importance of Normality Assumption in Using a T-Test: One Sample and Two Sample Emerging Trends in SocS15 Chennai Symposium