On Consistency and Limitation of Independent t-test Kolmogorov Smirnov Test and Mann Whitney U test

Usman, M.

Department of Statistics Federal Polytechnic Bali, PMB 05, Bali, Taraba State, Nigeria

Abstract: *It is worth repeating that if data are approximately normally distributed then parametric tests (as in the modules on hypothesis testing) are more appropriate. However, there are situations in which assumptions for a parametric test are violated and a nonparametric test is more appropriate**.** This study was aimed to investigates the strength and Limitation of independent t-test, Mann Whitney U test and Kolmogorov Smirnov test procedures on independent samples from unrelated population, under situations where the basic assumptions of parametric are not met for different sample size. Testing hypothesis on equality of means require assumptions to be made about the format of the data to be employed. Sometimes the test may depend on the assumption that a sample comes from a distribution in a particular family; if there is a doubt, then a non-parametric tests like Mann Whitney U test or Kolmogorov Smirnov test is employed. Random samples were simulated from Normal, Uniform, Exponential, Beta and Gamma distributions. The three tests procedures were applied on the simulated data sets at various sample sizes (small and moderate) and their Type I error and power of the test were studied in both situations under study.*
Keywords: *Independent Sample, Independent t-test, Mann Whitney U test, Kolmogorov Smirnov test,Type I error and Power of the test*

I.   Introduction

Nonparametric tests are usually less powerful than corresponding tests designed for use on data that come from a specific distribution. Thus, you are less likely to reject the null hypothesis when it is false. The tests often require one to modify the hypotheses. For example, most nonparametric tests about the population center are tests about the median instead of the mean. The test does not answer the same question as the corresponding parametric procedure ([1]). The strength of a nonparametric test resides in the fact that it can be applied without any assumption on the form of the underlying distribution. It is good for data with outliers and work well for ordinal data (data that have a defined order) because it based on ranks of data ([2]).

Nonparametric test is a hypothesis test that does not require the population's distribution to be characterized by certain parameters. For instance, many hypothesis tests rely on the assumption that the population follows a normal distribution with parameters μ and σ. Nonparametric tests do not have this assumption, so they are useful when your data are strongly nonnormal and resistant to transformation. However, nonparametric tests are not completely free of assumptions about your data, but it may also require the data to be an independent random sample. For example, salary data are heavily skewed to the right, with many people earning modest salaries and fewer people earning larger salaries. Therefore, one can use nonparametric tests to analyse such type of data ([3]).

In nonparametric tests very few assumptions are made about the distribution underlying the data and, in particular, it is not assumed to be a normal distribution. Some statisticians prefer to use the term distribution-free rather than nonparametric to describe these tests ([4]). In this article some hypothesis tests were developed in situations where the data come from a probability distribution whose underlying distribution is normal or not and different sample size were considered for each case of an independent sample. If the observations from two samples are unrelated, then we have independent observations. For example, a test could be done to investigate whether there is a difference in test anxiety based on educational level (i.e., your dependent variable would be "test anxiety" and your independent variable would be "educational level", which has two groups: "undergraduates" and "postgraduates"). Unrelated samples, also called unpaired samples or independent samples, are samples in which the cases in each sample are different. Often we are investigating differences in individuals, which means that when comparing two samples, an individual in one sample cannot also be a member of the other sample and vice versa. An example would be gender, an individual would have to be classified as either male or female not both.

Nonparametric statistical tests are concerned with the application of statistical data in nominal or ordinal scale to problems in pure science, social science, engineering and other related fields. Most of the present analysis carried out by non science and science oriented researchers are based on parametric test, and it is often reasonable to assume that observations come from a particular family of distributions. Moreover, experience backed by theory, suggest that for measurements, inferences based on the assumption that

observations form a random sample from some normal distribution may not be misleading even if the normality assumption is incorrect, but this is not always true. ([5]). Nonparametric tests often are used in conjunction with small samples, because for such samples the central limit theorem cannot be invoked. Nonparametric tests can be directed toward hypothesis concerning the form, dispersion or location (median) of the population. In the majority of the applications, the hypothesis is concerned with the value of a median, the difference between medians or the differences among several medians. This contrasts with the parametric procedures that are focused principally on population means. If normal model cannot be assumed for the data then the tests of hypothesis on means are not applicable. Nonparametric tests were created to overcome this difficulty. Nonparametric tests are often (but not always) based on the use of ranks; such as Kolmogorov Smirnov test, Wilcoxon rank test, Sign test, Mann Whitney U test, Kruskal Wallis test, etc ([6], [7]).

The objectives of this study are in two forms:

i.   To examine the effect of non-normality on parametric independent t-test and the nonparametric tests of the Mann Whitney U test and Kolmogorov Smirnov test effect.

ii. To examine the effects of sample size on the three test procedures based on type I error and power of test.

II. Materials And Methods

The materials used for the analysis were generated data using simulation procedures from selected continuous distributions. Since it is very difficult to get data that follows these distribution patterns, even if there is, it is very difficult to get the required number of replicates for the sample sizes of interest. The parametric (Independent t-test) and nonparametric (Mann Whitney U test and Kolmogorov Smirnov test), methods of analyzing independent sample were applied, to compare the performance of each test on the generated data from the Normal, Uniform, Exponential, Beta and Gamma distributions based on the underlying criteria for assessment

2.1 Simulation Procedures and Analysis

Random samples were simulated from Normal, Uniform, Exponential, Beta and Gammadistributions respectively for sample size of 5 10, 15, 25 and 30 which considered as small and moderate sample sizes respectively. Each test procedures were applied on the data sets at varying sample sizes and their Type I error and power of the tests were studied in each situation. At every replicate two samples were simulated independently from each distribution using the same parameters to form the required independent sample from the same population. The process was repeated 500 times for each independent sample size considered and results were displayed in table 1-5.

2.2 Criteria for Assessment and Test of Significance

Some decision must often be made between significance of a test or not. Turning the p-value into a binary decision allows us to examine two questions about the comparative value of statistical tests:

i. What percent of significant results will a researcher mistakenly judge to be in significant?

ii. What percent of reported significant results will actually be in significant?

Indeed the number of rejecting H_0 when it is true is counted for type I error and number of times H_0 is accepted when it is true was recorded as power of the test from each statistic under study.

Student's Independent t-test

The independent t-test, also called the two sample t-test or student's t-test, is an inferential statistical test that determines whether there is a statistically significant difference between the means in two unrelated samples on the same continuous, dependent variable. The test also asks whether a difference between two samples' averages is unlikely to have occurred because of random chance in sample selection. A difference is more likely to be meaningful and "real" if (i) the difference between the averages is large, (ii) the sample size is large, and (iii) responses are consistently close to the average values and not widely spread out (the standard deviation is low) ([8]).

Note that before performing any Independent t-Test the following assumptions must be satisfied:

i.   *Independence:* Observations within each sample must be independent (they don't influence each other)

ii. *Normal Distribution:* The scores in each population must be normally distributed

iii. *Homogeneity of Variance:* The two populations must have equal variances (the degree to which the distributions are spread out is approximately equal)

Let x_{1i} and x_{2i} be two of observation from different population of individual i, then we obtain

$$t = \frac{\bar{x}_1 - \bar{x}_2}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1 + n_2 - 1} \qquad (1)$$

where $\bar{x}_1$ is the 1st sample mean, $\bar{x}_2$ is the 2nd sample mean and $S_p^2 = \frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$ The null hypothesis (H_0) is reject if calculated t value is greater than table value (i.e. $|t|_{cal} > t_{\frac{\alpha}{2},n_1+n_2-2}$) otherwise we do not reject null hypothesis ([8]).

Mann Whitney U test

   Mann Whitney U-test is a nonparametric test alternative to the two independent sample t-test, which is called the Wilcoxon rank sum test. The test, sometimes called the Mann Whitney U Test or the Wilcoxon Rank Sum Test, is used to test whether two samples are likely to derive from the same population (i.e., that the two populations have the same shape). Some investigators interpret this test as comparing the medians between the two populations. The procedure for the test involves pooling the observations from the two samples into one combined sample and arrange the jointly values in an increasing order of magnitude and rank them in ascending order of magnitude, keeping track of which sample each observation comes from, [i.e. from 1 to $(n_1 + n_2)$], respectively. If there is ties, we assigned to each of the tied observations the mean of the ranks which they occupy. The test is based on W_1 the sum of the ranks of the values of first group or on W_2 the sum of the ranks of the values of second group. In actual practice, we base the test directly on W_1 and W_2, instead we use the related statistics

$$U_1 = W_1 - \frac{n_1(n_1 + 1)}{2} \qquad \text{or} \qquad U_2 = W_2 - \frac{n_2(n_2 + 1)}{2}$$

or the smaller of the two, which we denote by U i.e. $U = min(U_1, U_2)$

   This test is often performed as a two-sided test and, thus, the hypothesis indicates that the populations are not equal as opposed to specifying directionality. A one-sided hypothesis is used if interest lies in detecting a positive or negative shift in one population as compared to the other. Therefore we reject H_0 when $U \leq U_{\alpha/2}$ for two tailed test or when $U_2 \leq U_\alpha$ for one tailed test. Similar when $U_1 \geq U_\alpha$ we reject H_0 for one tailed test ([9]).

Kolmogorov-Smirnov Test

   Suppose there are differences we wish to detect than can manifest themselves in the location (mean, median), the spread (scale, variance), or perhaps even in the shape of the two distributions. It would be helpful to have a test to detect "general" difference. The Kolmogorov-Smirnov test is such a test, it tests $H0$: The distributions are the same against $H1$: The distributions are not the same ([10]). Let the two samples $X_1, X_2, ..., X_m$ and $Y_1, Y_2, ..., Y_n$ Denote $F1(x)$ the cumulative distribution function (CDF) for the distribution of the first population and $F2(y)$ the CDF for the second. Next, denote $\hat{F}_1$ the sample CDF for the first treatment and $\hat{F}_2$ be the sample CDF for the second treatment. The Kolmogorov-Smirnov statistic, denotes KS is found by

$$KS = max_v|\hat{F}_1(v) - \hat{F}_2(v)|$$

   The statistic is calculated by finding the maximum absolute value of the differences between the two independent sample CDFs. The hypothesis regarding the distributional form is rejected if the test statistic, KS, is greater than the critical value obtained from a table ([11]). These alternative formulations should be equivalent, but it is necessary to ensure that the test statistic is calculated in a way that is consistent with how the critical values were tabulated.

Table 1: Type I error (Normal Distribution)   α = 0.05

Sample Size(n)	Test Statistic		
	Independent T Test	Mann Whitney U Test	Kolmogorov Smirnov Test
5	0.0513	0.0487	0.0087
10	0.0373	0.0367	0.0093
15	0.0473	0.0480	0.0260
25	0.0540	0.0500	0.0407
30	0.0533	0.0513	0.0340
Averages	0.0486	0.0469	0.0237

Table 2: Type I error (Uniform Distribution) α = 0.05

Sample Size	Test Statistic		
	Independent T Test	Mann Whitney U Test	Kolmogorov Smirnov Test
5	0.0653	0.0607	0.0093
10	0.0627	0.0547	0.0147
15	0.0540	0.0447	0.0333
25	0.0433	0.0420	0.0440
30	0.0333	0.0333	0.0413
Averages	0.0517	0.0471	0.0285

Table 3: Type I error (Exponential Distribution)   α=0.05

Sample Size	Test Statistic		
	Independent T Test	Mann Whitney U Test	Kolmogorov Smirnov Test
5	0.0433	0.0580	0.0120
10	0.0427	0.0553	0.0153
15	0.0540	0.0413	0.0327
25	0.0360	0.0447	0.0440
30	0.0400	0.0340	0.0433
Averages	0.0432	0.0467	0.0295

Table 4: Type I error (Beta Distribution) α = 0.05

Sample Size	Test Statistic		
	Independent T Test	Mann Whitney U Test	Kolmogorov Smirnov Test
5	0.0553	0.0487	0.0100
10	0.0500	0.0473	0.0167
15	0.0467	0.0460	0.0280
25	0.0453	0.0433	0.0347
30	0.0467	0.0467	0.0367
Averages	0.0488	0.0464	0.0252

Table 5: Type I error (Gamma Distribution) α = 0.05

Sample Size	Test Statistic		
	Independent T Test	Mann Whitney U Test	Kolmogorov Smirnov Test
5	0.0387	0.0520	0.0087
10	0.0380	0.0480	0.0113
15	0.0487	0.0447	0.0260
25	0.0407	0.0460	0.0387
30	0.0440	0.0493	0.0380
Averages	0.0420	0.0480	0.0245

Table 6: Summary of the Type I error α = 0.05

Distributions	Test Statistic		
	Independent T Test	Mann Whitney U Test	Kolmogorov Smirnov Test
Normal	0.0486	0.0469	0.0237
Uniform	0.0517	0.0471	0.0285
Exponential	0.0432	0.0467	0.0295
Beta	0.0488	0.0464	0.0252
Gamma	0.0420	0.0480	0.0245

Fig 1: Potrayed the Type I Error for the three Tests
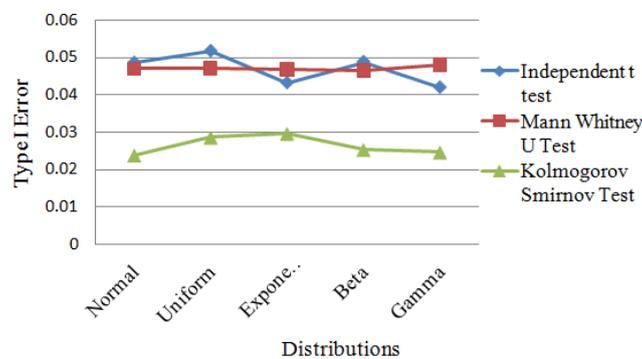
Table 7: Power of the Test (Normal Distribution)

Sample Size	Test Statistic		
	Independent T Test	Mann Whitney U Test	Kolmogorov Smirnov Test
5	0.7290	0.6130	0.5170
10	0.8020	0.7910	0.6650
15	0.8730	0.8680	0.7930
25	0.8960	0.8830	0.8580
30	0.9230	0.8976	0.8860
Averages	**0.8446**	**0.81052**	**0.7438**

Table 8: Power of the Test (Uniform Distribution)

Sample Size	Test Statistic		
	Independent T Test	*Mann Whitney U Test*	*Kolmogorov Smirnov Test*
5	0.8770	0.8915	0.8785
10	0.8955	0.8983	0.9051
15	0.9015	0.9414	0.9496
25	0.9347	0.9624	0.9732
30	0.9581	0.9782	0.9914
Averages	**0.91336**	**0.93436**	**0.93956**

Table 9: Power of the Test (Exponential Distribution)

Sample Size	Test Statistic		
	Independent T Test	*Mann Whitney U Test*	*Kolmogorov Smirnov Test*
5	0.1550	0.1120	0.1090
10	0.4570	0.4560	0.1910
15	0.6840	0.6160	0.3420
25	0.8040	0.7060	0.4430
30	0.8940	0.8780	0.5110
Averages	**0.5988**	**0.5536**	**0.3192**

Table 10: Power of the Test (Beta Distribution)

Sample Size	Test Statistic		
	Independent T Test	*Mann Whitney U Test*	*Kolmogorov Smirnov Test*
5	0.5230	0.6770	0.5190
10	0.8420	0.8580	0.6480
15	0.9320	0.9370	0.8170
25	0.9450	0.9501	0.8780
30	0.9510	0.9648	0.9150
Averages	**0.8386**	**0.87738**	**0.7554**

Table 11: Power of the Test (Gamma Distribution)

Sample Size	Test Statistic		
	Independent T Test	*Mann Whitney U Test*	*Kolmogorov Smirnov Test*
5	0.6760	0.6090	0.6140
10	0.8070	0.7960	0.6450
15	0.8838	0.8880	0.8105
25	0.9160	0.9065	0.8880
30	0.9317	0.9245	0.9160
Averages	**0.8429**	**0.8248**	**0.7747**

Table 12: Summary of the Power of the Tests

Distributions	Test Statistic		
	Independent T Test	*Mann Whitney U Test*	*Kolmogorov Smirnov Test*
Normal	0.8446	0.8105	0.7438
Uniform	0.9134	0.9344	0.9396
Exponential	0.5988	0.5536	0.3192
Beta	0.8386	0.8774	0.7554
Gamma	0.8429	0.8248	0.7747

Fig 2: Shows the Power of the Three Tests
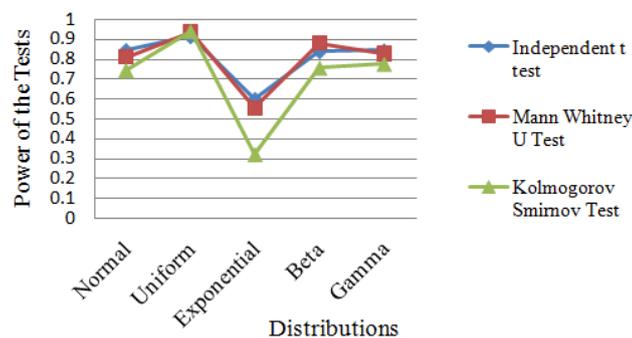
III.   Discussion Of Results

Tables 1 – 5 and 7 – 11 indicate results of analyses using the independent t-test, Mann Whitney U test and Kolmogorov Smirnov test on how the tests perform based on the type I error and power of the test, both being compared at the 5% level of significance for two tailed test in each case. The average of each value of the type I error and power of the test were calculated and recorded under each statistical test for easy comparison. Figure 6 and 12 displayed the results of the analysis on how the tests perform based on the type I error and power of the test.

The type I error of the independent t-test, Mann Whitney U test and Kolmogorov Smirnov test increases from normal distributions and started to decrease at sample size of 30 for independent t test and Kolmogorov Smirnov test as we can see in the table 1. However, the Kolmogorov Smirnov test increases in the type I error from the uniform, exponential and beta distributions, while independent t test and Mann Whitney U test type I errors decreases as the sample size increase for the respective distribution. Except that independent t test type I error value fluctuate from the exponential distribution as can be seen in table 3. For data generated from gamma distribution, the type I error for independent t test fluctuate as well. However, the Kolmogorov Smirnov test increases in the type I error for sample size 5 to 25, while Mann Whitney U test type I errors decreases as the sample size increase.

The Mann Whitney U test has the type I error that is closest to α from exponential and gamma distribution while the independent t-test has the type I error that is closest to α from normal, uniform and beta distributions. The Kolmogorov Smirnov tests has the lowest type I error from all distributions consider for this study (see table 6). The power of the independent t-test as well as that of nonparametric tests (Mann Whitney U test and Kolmogorov Smirnov tests) increases as the sample size increases for data generated from the five distributions consider for this study with values range from 0.6130 to 0.9230, 0.8770 to 0.9914, 0.1090 to 0.8940, 0.5190 to 0.9648 and 0.6090 to 0.9317 for normal, uniform, exponential, beta and gamma distributions from sample size of 5 to 30 respectively. The t-test test has the highest power from the data generated from normal, exponential and gamma distributions as shown in table 12 followed by Mann Whitney U test. In the data generated from uniform distribution Kolmogorov Smirnov test has the highest power followed by Mann Whitney U test, while the Mann Whitney U test has the highest power from beta distribution followed by the Independent t test. These can be seen clearly in the figure 2 presented above for the summary of the power of the tests.

IV.   Conclusion

It was observed that the Mann Whitney U test has the highest power of the test from the data generated from beta distribution and therefore consider as the most powerful test in that respect while Kolmogorov Smirnov test has the highest power of the test from the data generated from uniform distribution as shown in table 12 and considered as the most powerful test for that distribution. Meanwhile, the independent t test was consider as the most powerful test from the data generated from normal, exponential and gamma distributions having the highest power of the test in that respect. However, there is no significant differences in the power of the tests when rounded to two decimal places, if they are compared based on the simulated data from the five selected distributions, using small sample sizes at the 5% levels of significance. Hence, the independent t-test is the most suitable test when the underline distribution is normal and when sample sizes are large for any distributions as reported in the table 1 – 5 and 7 – 11. However the two nonparametric tests are indeed alternative tests to t-test when the assumption of normality is not met for independent sample.

References

[1].   J. D. Gibbons and S. Chakraborti, *Nonparametric Statistical Inference*, 4th edition. CRC   Press. ISBN 0-8247-4052-1. 2003
[2].   G. W. Corder and D. I. Foreman, *Nonparametric Statistics: A Step-by-Step Approach.* John Wiley & Sons New York:. ISBN 978-1118840313. 2014
[3].   T. P. Hettmansperger and J. W. McKean, *Robust Nonparametric Statistical Methods.* Kendall's Library of Statistics 5 (First edition). London: Edward Arnold. New York: John Wiley & Sons. ISBN 0-340-54937-8.MR 1604954. 1998
[4].   G. M. Clarke and D. Cooke, *A Basic Course in Statistics* 4th Edition Nicki Dennis London NW1 3BH. 1998
[5].   P. Sprent, *Applied Non-parametric Statistical Methods*. 2nd Edition. Chapman and Hall 1993
[6].   M. Usman and A. I. Maksha, An Efficient Alternative to t-Test For One Sample Nonnormal Data, *Journal of Applied Science and Technology*, Auchi Polytechnic,   Nigeria. 2010
[7].   I. Akeyede and S. G. Akinyemi, Power Comparison of Sign and Wilcoxon Sign Rank Test Under Non Normal, *African Journal of Physical Sciences,* Devon Science Publication. 2010
[8].   B. A. Oyejola and S. B. Adebayo, *Basic Statistics for Biology and Agriculture Students,* Olad Publishers, Ilorin, Nigeria. 2004
[9].   S. Siegel and N. J. Castellan, *Non-parametric Statistics for the Behavioural Science,* 2nd   Edition New York: McGraw Hill. 1988
[10].   A. Stuart, K. Ord and S. Arnold, *Classical Inference and the Linear Model.* Kendall's Advanced Theory of Statistics, 2A 6th edition. London: Arnold. ISBN 0-340-66230-1.MR 1687411. 1999
[11].   W. W. Daniel, *"Kolmogorov–Smirnov one-sample test".* Applied Nonparametric Statistics, 2nd edition. Boston: PWS-Kent. ISBN 0-534-91976-6. 1990