International Journal of Mathematics and Statistics Invention (IJMSI)
E-ISSN: 2321- 4767 P-ISSN: 2321-4759
www.ijmsi.org Volume 9 Issue 1 || January, 2021 || PP 01-11

Evaluation of Two-sample T-test and Analogous Nonparametric Tests: A Simulation Approach

Khairul Islam* and Tanweer Shapla
Department of Mathematics and Statistics
Eastern Michigan University, Ypsilanti, Michigan 48197 USA

ABSTRACT**:** *A two-sample t-test is the most powerful test for comparing two population means under normal models. In real-life, the assumption of normality may not meet. Under these circumstances, a transformed two-sample t-test or alternately, analogous non-parametric tests such as Mann-Whitney test and Kolmogorov Smirnov tests may be employed to achieve certain objectives. To recommend the best test under non-normal models, it is imperative to evaluate performance of underlying tests using Type I error probability and power of the test via Monte Carlo simulation at various sample sizes. In this study, we simulate independent samples from skewed distributions with varying levels of skewness, along with symmetric distributions to evaluate performance of four underlying tests, namely, two-sample t-test, transformed two-sample test, Mann-Whitney test and Kolmogorov Smirnov test.*

KEYWORDS**:** *Skewed distribution, Transformed t-test, Mann-Whitney test, Kolmogorov-Smirnov test, Power of the test.*

Date of Submission: 15-01-2021 Date of Acceptance: 31-01-2021

I. INTRODUCTION

Given two independent samples, we often wish to test if the two populations the samples come from have the same mean. The test requires the normality of the two population distributions. One may use a two-sample z-test or a t-test if the two population variances are known or unknown but equal ([1-2]). In real life, it is very unlikely that the two population variances are known. For the present study, we therefore, ignore the case where population variances might be known, and undertake the case with unknown but equal variance scenario. Given these circumstances, a two-sample t-test is the most powerful test, which appears in any standard statistics book [1-3] and is briefly presented in the method section.

One adverse reality in using the t-test might be the non-normality of the two population distributions. In the violation of the normality assumptions, the t-test is vulnerable in that it is not robust. An alternative and the most popular analogue of the t-test is the Wilcoxon and Mann Whitney (WMW) test [4-10], which does not require the normality assumption. Another NP test in the two-sample situation is the Kolmogorov-Smirnov (KS) test, which can also be applied in the scenarios of the WMW test. In a recent study [11], the consistency and limitation of t-test, KS test and WMW test are studied. Because parametric tests are more powerful than any nonparametric test, practitioners with preference to the parametric tests often opt for transformation of non-normal data with suitable transformations ([12]-[16]).

In this paper, we consider a transformed t-test obtained via Box-Cox transformation ([12]) with a goodness of fit technique. To support the use of transformed t-test in the comparison, it is worth mentioning that the nonparametric WMW and KS tests are transformed tests which replace original data values by ranks or signs. Therefore, indeed, WMW and KS tests are a particular form of transformed tests utilized in the violation of the normality. If the WMW test or KS test could be compared with t-test ([11]), it makes sense to include a transformed t-test in the comparison. However, unlike ([11]), we investigate the effect of varying degree of skewness or mean differences in the two populations so as to compare performance of the t-test following transformation or without transformation with those of the NP tests.

The main objectives of this study are as follows:
(i) To examine the effect of non-normality on parametric two-sample t-test, transformed t-test, and the nonparametric analogous Mann-Whitney U test and Kolmogorov Smirnov test.
(ii) To examine the effect of varying sample sizes, both balanced (i.e., $m = n$) and unbalanced (i.e., $m \neq n$) cases, on the four underlying tests.
(iii) To examine the effect or sensitivity of varying degree of skewness or mean differences on the four underlying tests.

The effect of the underlying tests in reference to the specified objectives (i)-(iii) will be assessed in terms of the Type I error rate and power via simulation studies.

II. METHODS

Given two independent samples $X_1, X_2, \cdots, X_m$ and $Y_1, Y_2, \cdots, Y_n$ from two populations X and Y with unknown means μ_x and μ_y, and common but unknown variance, say σ^2, we wish to test $H_{01}: \mu_x = \mu_y$. Under the assumption of the normality of the distributions of X and Y, the test of H_{01} is called a two-sample t-test. Note that if the distributions of X and Y deviate from normality, the t-test is invalid. Then, under the assumptions that X and Y have the continuous cdfs F_X and F_Y, respectively, one might test

$$H_{02}: F_X(x) = F_Y(x) \text{ for all } x$$

via the nonparametric Wilcoxon and Mann Whitney (WMW) test or the Kolmogorov-Smirnov (KS) test. The test of H_{02} is called the test of the identical distribution of X and Y. Note that if the distributions of X and Y are identical, they will have the same location parameters, means or medians, say μ_x and μ_y. Therefore, the test of H_{02} is a more general test of the two-sample problems, and can be thought of a particular form of the non-parametric analogue of the t-test for $H_{01}: \mu_x = \mu_y$ with the unknown but equal variance scenario.

As such, in this paper, we evaluate performance of the four tests– t of $H_{01}: \mu_x = \mu_y$, Wilcoxon and Mann-Whitney (WMW) and Kolmogorov-Smirnov (KS) tests of $H_{02}: F_X(x) = F_Y(x)$, and transformed t test of $H_{01}: \mu_x = \mu_y$ for two indpendent sample situations.

T-TEST

Let $s_p^2 = \frac{(m-1)s_x^2 + (n-1)s_y^2}{m+n-2}$ be the pooled estimate of the common unknown variance σ^2, where s_x^2 and s_y^2 are sample variances of the two given samples $X_1, X_2, \cdots, X_m$ and $Y_1, Y_2, \cdots, Y_n$. Then, the test of $H_0: \mu_x = \mu_y$, commonly known as the pooled t-test, is defined by

$$T = \frac{\overline{X} - \overline{Y}}{s_p \times \sqrt{\frac{1}{m} + \frac{1}{n}}}$$

Under H_0, the test statistic T follows Student's t-distribution with m+n-2 degrees of freedom. This test is valid as long as the two given samples $X_1, X_2, \cdots, X_m$ and $Y_1, Y_2, \cdots, Y_n$ come from two normal populations with unknown but a common variance.

In the violation of the normality assumptions, the most common practice is to employ the WMW or KS test.

WILCOXON AND MANN WHITNEY TEST

Given two independent samples $X_1, X_2, \cdots, X_m$ and $Y_1, Y_2, \cdots, Y_n$ from two populations X and Y with continuous cdfs F_X and F_Y, respectively, the Wilcoxon and Mann Whiteney (WMW) test is specified by

$$H_{02}: F_X(x) = F_Y(x) \text{ for all } x$$

against the alternative

$$H_{12}: F_Y(x) = F_X(x - \theta) \text{ for all } x \text{ and some } \theta \neq 0.$$

The alternative H_{12} is called the location alternative. Under the H_{02}, if we make assumptions concerning the form of the common population, in particular that it is normal with a common but unknown variance, then the test of H_{02} boils down to the the non-parametric analogue of the t-test for $H_{01}: \mu_x = \mu_y$.

Because, in general, the test of H_{02} does not make any assumption of the form of the distributions, the test is known as a more general test of the two-sample situation. Note that the cumulative distribution function of the population Y under H_{12} is the same as that of the the population X but shifted by $\theta \neq 0$. If $\theta < 0$, the median of the population X is larger than the median of the population Y, and vice versa if $\theta > 0$.

The Wilcoxon test statistic of H_{02} is given by

$$W_N = \sum_{i=1}^{N} i\, Z_i$$

where Z_i is the indicator random variable defined by $Z = (Z_1, Z_2, \cdots, Z_N)$ with $N = m + n$ and $Z_i = 1$ if the ith random variable in the combined ordered sample is an X and $Z_i = 0$ if it is a Y, for $i = 1,2, \cdots, N$. It follows that the rank of the observation for which Z_i is an indicator is i, and therefore the vector Z indicates the rank-order statistics of the combined samples and in addition identifies the sample to which each observation belongs.

If there is no ties, under null hypothesis, the mean and variance of W_N are given by $E[W_N] = \frac{m(N+1)}{2}$ and $V[W_N] = \frac{mn(N+1)}{12}$, respectively. Given mean and variance of the test statistic, one could employ a normal approximation to the rank sum test.

The Wilcoxon test of H_{02} is linearly related to another test due to Mann-Whitney derived by comparing each $X_i, i = 1,2,\cdots,m$ with each $Y_j, j = 1,2,\cdots,n$ and is given by

$$U = \sum_{i=1}^{m}\sum_{j=1}^{n} D_{i,j}$$

where

$$D_{i,j} = \begin{cases} 1, \text{if } Y_j < X_i \\ 0, \text{if } Y_j > X_i \end{cases}$$

It follows that $U = W_N - \frac{m(m+1)}{2}$, and due to the fact, the test is most popularly known as the Wilcoxon and Mann Whiteney (WMW). More details about the test is available in Gibbons [4-5]. In this paper, we implement the function wilcox.test() available in R for all computation and simulation.

KOLMOGOROV-SMIRNOV TEST

Let $X_1, X_2, \ldots, X_m$ and $Y_1, Y_2, \ldots, Y_n$ be two independent samples from populations with unknown CDFs F_X and F_Y. Let $S_m(x)$ and $T_n(x)$ be corresponding empirical distribution functions (EDFs). The Kolmogorov-Smirnov (KS) test statistic of

$$H_{02}: F_X(x) = F_Y(x) \text{ for all } x$$

against the alternative

$$H_1: F_Y(x) \neq F_Y(x) \text{ for some } x$$

is given by

$$D_{m,n} = \sup_x |S_m(x) - T_n(x)|$$

For large samples, the null hypothesis is rejected at level α if $D_{m,n} > D_{cv}$, where D_{cv} is the critical value given by $D_{cv} = c(\alpha) \times \sqrt{\frac{n+m}{nm}}$ with the constant $c(\alpha)$ defined by $c(\alpha) = \sqrt{-\frac{1}{2} \times \ln(\alpha)}$. One can use the critical value of the test statistic KS using the table ([17]). In this paper, we utilize the function ks.test() available in R.

BOX-COX TRANSFORMED T-TEST

Researchers with preference to parametric tests try to transform non-normal data to normality or nearing normality ([12]) so as to enjoy the parametric test. For example, given two samples with non-negative values $X_1, \cdots, X_m$ and $Y_1, \cdots, Y_n$ from positively skewed distributions, and a scalar λ, one might transform X_i by

$$X_i(\lambda) = \begin{cases} (X_i^\lambda - 1)/\lambda, \text{ if } \lambda > 0 \\ \log(X_i), \text{ if } \lambda = 0 \end{cases}$$

and the transformation of Y_j to $Y_j(\lambda)$ is defined in a similar manner.

Given the transformation, let $S^2(\lambda)$ be the pooled maximum likelihood estimate (MLE) of the variance of the transformed data given by $s^2(\lambda) = \frac{\sum_{i=1}^{m}(X_i(\lambda) - \overline{X}(\lambda))^2 + \sum_{j=1}^{n}(Y_j(\lambda) - \overline{Y}(\lambda))^2}{m+n}$ where $\overline{X}(\lambda)$ and $\overline{Y}(\lambda)$ are sample variance of transformed X and Y samples. Given the transformation to normality or nearing normality, one can utilize the profiled log-likelihood function

$$l(\lambda) = -\{(m+n)/2\}\log(s^2(\lambda)) + \lambda\left\{\sum_{i=1}^{m}\log(X_i) + \sum_{j=1}^{n}\log(Y_j)\right\}$$

to estimate λ by the MLE $\hat{\lambda}_l$ over a pre-specified set I of values of λ [12] so that $l(\lambda)$ is maximized. Then, the transformed t-test given by

$$T(\hat{\lambda}_l) = \frac{\overline{X}(\hat{\lambda}_l) - \overline{Y}(\hat{\lambda}_l)}{s(\hat{\lambda}_l)\sqrt{1/m + 1/n}}$$

can be used to test $H_{01}: \mu_x = \mu_y$ by comparing $T(\hat{\lambda}_l)$ with the critical value $t_{\alpha/2; m+n-2)}$.

Hinkley (1975), Hernandez and Johnson (1980), Bickel and Doksum (1981), and many others investigated the asymptotic properties of the parameter estimates. Chen and Loh (1992) and Chen (1995) proved that the Box-

Cox transformed test is typically more efficient asymptotically than the t-test without transformation. Islam and Chen (2007) justified the use of transformed t-test by fitting a t distribution to transformed data.

Alternative to MLE method to estimate λ, Islam and Shapla (2015) proposed to estimate λ via the univariate goodness of fit. They argued via simulation and example that transformed test via univariate goodness of fit performs better than that of MLE method. In this paper, we compare performance of Box-Cox type transformed t-test with a simpler log-transformed t-test, along with other underlying non-parameteric tests.

LOG-TRANSFORMED TWO-SAMPLE T-TEST

Due to the simplicity of implementation, log-transformation is widely used for transforming substantially skewed data to reduce skewness. The log-transformation of x is defined as follows:

(a) $x^* = \log(x)$, where x is substantially positive skewed.

(b) $x^* = \log(x + c)$, where x is substantially positive skewed with existence of zero values; c is a constant added to each data value so that log is defined.

(c) $x^* = \log(K - x)$, where x is negatively substantially skewed; $K = \max(x) + 1$.

Since beta, exponential and gamma distribution are skewed, it is expected that the log-transformed two-sample t-test will be the most important alternative to the Box-Cox transformed t-test via MLE method or univariate goodness of fit or other non-parametric tests in the violation of the normality, which we investigate via simulation and example.

III. SIMULATION

We simulate two samples from three different distributions, namely, beta, exponential and gamma distribution. The population distributions, the simulation will be carried out are subject to the following characteristics.

(1) We simulate X and Y from $Beta(\alpha,\beta)$ distribution. Note that if $X \sim Beta(\alpha,\beta)$, then $E[X] = \frac{\alpha}{\alpha+\beta}$ and $Skewness = \frac{2\times(\beta-\alpha)\times\sqrt{\alpha+\beta+1}}{(\alpha+\beta+2)\times\sqrt{\alpha\times\beta}}$. Given these facts, we arbitrarily choose $\alpha = 1$ and $\beta = 3$ so that with mean=0.25 and skewness=0.86. The shift of mean θ is considered as in the set {0, 0.05, 0.1, 0.12, 0.15} so that the power of the test or estimated significance levels is away from 0 and 1. If $\theta = 0$, the two samples X and Y come from the null model.

(2) We consider $X, Y \sim Exp(\alpha)$ with $\alpha = 1$ chosen arbitraily so that mean=1 and skewness=2. The shift of mean θ is considered as in the set {0, 0.2, 0.4, 0.5, 0.6} so that the power of the test or estimated significance levels is away from 0 and 1. If $\theta = 0$, the two samples X and Y come from the null model.

(3) The parameters of gamma $G(\alpha,\beta)$ distributions are chosen arbitrarily to allow varying levels of skewnesses in the population distributions. The values of the shape parameter α is considered from the set {4, 1, 0.25, 0.0625} with the corresponding β from the set {0.25, 1, 4, 16} so as to fix mean at 1 and skewness at {1, 2, 4, 8}.

(4) In all simulation, the Monte Carlo size is considered to 10,000. The estimated power of an underlying test is the proportion of rejection of the implementation of the test over 10,000 samples. Note that the location shift of $\theta = 0$ refers to the simulation under the null model, and the proportion of rejection under the null model over 10,000 simulations corresponds to the estimated level of significance at 5% level of significance.

TABLE 1.1: Simulated significance level and power of tests for $X \sim Beta(1,3) + \theta$, $Y \sim Beta(1,3)$

m,n	t	W	KS	$log-tt$	$bc-ttL$	$bc-ttGoF$
			θ=0			
15,10	0.051	0.047	0.030	0.047	0.054	0.058
15,15	0.052	0.045	0.027	0.048	0.053	0.060
20,15	0.050	0.047	0.037	0.049	0.053	0.057
20,20	0.049	0.048	0.035	0.049	0.051	0.054
25,20	0.051	0.048	0.036	0.050	0.054	0.056
25,25	0.049	0.046	0.034	0.047	0.050	0.052
30,25	0.046	0.044	0.028	0.047	0.048	0.050
			$\theta = 0.05$			
15,10	0.093	0.108	0.058	**0.164**	0.137	0.150
15,15	0.102	0.111	0.058	**0.156**	0.139	0.148

20,15	0.122	0.136	0.087	**0.214**	0.175	0.190
20,20	0.127	0.144	0.081	**0.216**	0.179	0.190
25,20	0.131	0.151	0.093	**0.260**	0.197	0.214
25,25	0.147	0.171	0.105	**0.278**	0.220	0.233
30,25	0.153	0.183	0.103	**0.323**	0.242	0.260
			$\theta = 0.10$			
15,10	0.232	0.263	0.161	**0.393**	0.333	0.376
15,15	0.279	0.309	0.182	**0.431**	0.381	0.416
20,15	0.307	0.345	0.265	**0.533**	0.441	0.494
20,20	0.351	0.407	0.289	**0.565**	0.483	0.524
25,20	0.398	0.445	0.336	**0.656**	0.543	0.600
25,25	0.434	0.487	0.378	**0.691**	0.587	0.640
30,25	0.472	0.531	0.392	**0.755**	0.638	0.693
			$\theta = 0.12$			
15,10	0.319	0.356	0.228	**0.510**	0.439	0.494
15,15	0.364	0.396	0.260	**0.541**	0.477	0.523
20,15	0.429	0.467	0.387	**0.667**	0.573	0.636
20,20	0.478	0.535	0.418	**0.704**	0.617	0.671
25,20	0.530	0.574	0.464	**0.786**	0.682	0.744
25,25	0.579	0.629	0.546	**0.825**	0.728	0.781
30,25	0.624	0.673	0.564	**0.869**	0.769	0.828
			$\theta = 0.15$			
15,10	0.458	0.480	0.338	**0.655**	0.581	0.648
15,15	0.535	0.559	0.412	**0.710**	0.654	0.705
20,15	0.600	0.626	0.565	**0.811**	0.727	0.791
20,20	0.665	0.704	0.623	**0.854**	0.783	0.834
25,20	0.715	0.744	0.659	**0.904**	0.831	0.880
25,25	0.765	0.796	0.761	**0.932**	0.873	0.908
30,25	0.798	0.828	0.761	**0.957**	0.899	0.937

TABLE 1.2: Simulated significance level and power of tests for $X{\sim}Exp(1)+\theta$, $Y{\sim}Exp(1)$

m,n	t	W	KS	$log-tt$	$bc-ttL$	$bc-ttGoF$
			θ=0			
15,10	0.043	0.047	0.030	0.047	0.051	0.057
15,15	0.047	0.049	0.028	0.052	0.057	0.061
20,15	0.047	0.048	0.035	0.051	0.055	0.059
20,20	0.051	0.052	0.034	0.052	0.055	0.059
25,20	0.049	0.049	0.037	0.050	0.053	0.056
25,25	0.050	0.047	0.036	0.049	0.052	0.054
30,25	0.048	0.048	0.034	0.051	0.052	0.054
			$\theta = 0.2$			
15,10	0.078	0.123	0.072	**0.183**	0.157	0.182
15,15	0.082	0.119	0.064	**0.173**	0.153	0.170
20,15	0.094	0.147	0.107	**0.230**	0.192	0.219
20,20	0.100	0.160	0.098	**0.239**	0.198	0.220

25,20	0.109	0.175	0.122	**0.291**	0.232	0.265
25,25	0.115	0.184	0.130	**0.300**	0.237	0.267
30,25	0.123	0.219	0.139	**0.359**	0.285	0.323
			$\theta = 0.4$			
15,10	0.185	0.283	0.203	**0.414**	0.355	0.424
15,15	0.216	0.329	0.242	**0.451**	0.391	0.451
20,15	0.235	0.382	0.361	**0.558**	0.470	0.555
20,20	0.258	0.428	0.392	**0.588**	0.490	0.574
25,20	0.290	0.476	0.439	**0.679**	0.570	0.663
25,25	0.313	0.529	0.519	**0.711**	0.607	0.684
30,25	0.336	0.558	0.528	**0.766**	0.652	0.744
			$\theta = 0.5$			
15,10	0.257	0.386	0.303	0.540	0.467	**0.561**
15,15	0.303	0.451	0.377	0.591	0.520	**0.598**
20,15	0.334	0.508	0.536	0.695	0.597	**0.702**
20,20	0.369	0.590	0.584	**0.749**	0.650	0.739
25,20	0.411	0.622	0.619	**0.813**	0.710	0.808
25,25	0.444	0.685	0.724	**0.850**	0.751	0.836
30,25	0.477	0.722	0.740	**0.889**	0.797	0.877
			$\theta = 0.6$			
15,10	0.344	0.485	0.425	0.655	0.572	**0.678**
15,15	0.399	0.567	0.533	0.710	0.630	**0.718**
20,15	0.445	0.629	0.688	0.807	0.711	**0.817**
20,20	0.485	0.714	0.741	**0.849**	0.764	0.845
25,20	0.540	0.753	0.777	**0.902**	0.818	0.897
25,25	0.574	0.796	0.865	**0.924**	0.850	0.915
30,25	0.608	0.838	0.877	**0.951**	0.889	0.947

TABLE 1.3: Simulated level of significance and power of tests for gamma distributions

			skewness=1			
m, n	t	W	KS	$log-tt$	$bc-ttL$	$bc-ttGoF$
			$\theta = 0$			
15,10	0.048	0.047	0.030	0.048	0.050	0.056
15,15	0.049	0.045	0.025	0.050	0.054	0.059
20,15	0.050	0.046	0.033	0.050	0.052	0.056
20,20	0.052	0.050	0.034	0.053	0.054	0.058
25,20	0.051	0.049	0.037	0.048	0.052	0.055
25,25	0.051	0.049	0.034	0.051	0.053	0.056
30,25	0.051	0.047	0.031	0.049	0.050	0.052
			$\theta = 0.1$			
15,10	0.071	0.076	0.048	0.089	0.088	**0.096**
15,15	0.082	0.078	0.047	0.092	0.092	**0.100**
20,15	0.085	0.085	0.057	0.102	0.099	**0.105**
20,20	0.095	0.099	0.061	0.111	0.108	**0.114**
25,20	0.103	0.106	0.075	0.126	0.120	**0.127**
25,25	0.110	0.110	0.073	0.128	0.123	**0.129**

30,25	0.112	0.123	0.077	**0.146**	0.139	0.145
			$\theta = 0.2$			
15,10	0.157	0.167	0.101	0.208	0.197	**0.212**
15,15	0.189	0.190	0.104	0.229	0.221	**0.236**
20,15	0.218	0.227	0.157	**0.279**	0.264	**0.279**
20,20	0.243	0.260	0.166	**0.303**	0.290	0.303
25,20	0.270	0.297	0.197	**0.355**	0.334	0.349
25,25	0.287	0.306	0.216	**0.372**	0.351	0.364
30,25	0.311	0.345	0.234	**0.417**	0.390	0.407
			$\theta = 0.3$			
15,10	0.305	0.321	0.206	0.391	0.371	**0.395**
15,15	0.376	0.382	0.245	0.453	0.437	**0.458**
20,15	0.402	0.426	0.319	**0.518**	0.487	0.513
20,20	0.465	0.501	0.359	**0.569**	0.542	0.565
25,20	0.511	0.547	0.410	**0.638**	0.605	0.628
25,25	0.552	0.590	0.459	**0.673**	0.641	0.661
30,25	0.594	0.646	0.486	**0.734**	0.696	0.717
			$\theta = 0.4$			
15,10	0.492	0.510	0.347	0.600	0.571	**0.604**
15,15	0.573	0.595	0.417	0.676	0.652	**0.678**
20,15	0.626	0.652	0.537	**0.751**	0.716	0.748
20,20	0.707	0.747	0.607	**0.813**	0.785	0.808
25,20	0.745	0.781	0.657	**0.858**	0.827	0.850
25,25	0.790	0.826	0.717	**0.891**	0.862	0.880
30,25	0.820	0.852	0.748	**0.915**	0.889	0.907

TABLE 1.3: continued

			skewness=2			
m, n	t	W	KS	$log - tt$	$bc - ttL$	$bc - ttGoF$
			$\theta = 0$			
15,10	0.044	0.048	0.028	0.049	0.054	0.059
15,15	0.046	0.046	0.027	0.050	0.054	0.060
20,15	0.049	0.047	0.033	0.052	0.054	0.059
20,20	0.044	0.047	0.034	0.045	0.047	0.051
25,20	0.050	0.052	0.040	0.053	0.055	0.058
25,25	0.046	0.045	0.035	0.050	0.050	0.053
30,25	0.047	0.047	0.034	0.046	0.048	0.050
			$\theta = 0.1$			
15,10	0.045	0.064	0.039	**0.092**	0.082	0.091
15,15	0.055	0.069	0.037	**0.090**	0.084	0.090
20,15	0.054	0.071	0.047	**0.108**	0.090	0.100
20,20	0.053	0.073	0.045	**0.101**	0.088	0.094
25,20	0.063	0.083	0.052	**0.127**	0.106	0.113
25,25	0.064	0.083	0.051	**0.128**	0.107	0.114
30,25	0.064	0.091	0.054	**0.148**	0.118	0.129
			$\theta = 0.2$			
15,10	0.072	0.118	0.069	**0.176**	0.153	**0.176**

15,15	0.086	0.124	0.066	**0.180**	0.159	0.179
20,15	0.089	0.143	0.103	**0.225**	0.186	0.212
20,20	0.105	0.167	0.108	**0.246**	0.204	0.228
25,20	0.106	0.177	0.118	**0.294**	0.234	0.268
25,25	0.116	0.186	0.128	**0.299**	0.239	0.268
30,25	0.127	0.218	0.143	**0.355**	0.285	0.319
			$\theta = 0.3$			
15,10	0.119	0.203	0.128	0.299	0.255	**0.304**
15,15	0.141	0.223	0.136	**0.313**	0.273	0.311
20,15	0.151	0.251	0.214	**0.394**	0.325	0.385
20,20	0.168	0.294	0.227	**0.426**	0.351	0.401
25,20	0.185	0.323	0.251	**0.495**	0.402	0.471
25,25	0.200	0.347	0.302	**0.520**	0.420	0.485
30,25	0.214	0.390	0.306	**0.583**	0.475	0.550
			$\theta = 0.4$			
15,10	0.189	0.290	0.210	0.423	0.363	**0.435**
15,15	0.216	0.334	0.239	0.455	0.395	**0.456**
20,15	0.240	0.378	0.366	**0.557**	0.467	0.554
20,20	0.265	0.448	0.396	**0.606**	0.512	0.589
25,20	0.280	0.477	0.437	**0.677**	0.567	0.661
25,25	0.318	0.526	0.524	**0.711**	0.602	0.686
30,25	0.335	0.562	0.535	**0.768**	0.647	0.747

TABLE 1.3: continued

			skewness=4			
m, n	t	W	KS	$log - tt$	$bc - ttL$	$bc - ttGoF$
			$\theta = 0$			
15,10	0.034	0.049	0.031	0.047	0.055	0.061
15,15	0.033	0.046	0.026	0.046	0.052	0.056
20,15	0.036	0.048	0.034	0.047	0.053	0.057
20,20	0.036	0.048	0.035	0.048	0.050	0.053
25,20	0.036	0.047	0.036	0.049	0.051	0.054
25,25	0.041	0.051	0.035	0.050	0.056	0.057
30,25	0.041	0.044	0.031	0.042	0.045	0.047
			$\theta = 0.1$			
15,10	0.034	0.179	0.247	0.349	0.249	**0.382**
15,15	0.037	0.187	0.313	0.341	0.249	**0.354**
20,15	0.040	0.232	**0.517**	0.472	0.316	0.493
20,20	0.043	0.248	**0.538**	0.484	0.324	0.472
25,20	0.040	0.274	0.553	**0.588**	0.375	0.574
25,25	0.049	0.304	**0.719**	0.610	0.392	0.568
30,25	0.049	0.327	**0.715**	0.688	0.438	0.648
			$\theta = 0.2$			
15,10	0.045	0.284	0.432	0.511	0.381	**0.555**
15,15	0.061	0.320	**0.559**	0.534	0.407	0.555
20,15	0.052	0.371	**0.748**	0.666	0.483	0.692
20,20	0.063	0.427	**0.802**	0.700	0.516	0.696

25,20	0.060	0.456	**0.797**	0.790	0.575	0.784
25,25	0.069	0.503	**0.920**	0.824	0.605	0.794
30,25	0.062	0.524	**0.913**	0.876	0.647	0.851
			$\theta = 0.3$			
15,10	0.066	0.372	0.560	0.614	0.480	**0.644**
15,15	0.080	0.429	**0.709**	0.652	0.522	**0.680**
20,15	0.076	0.485	**0.857**	0.777	0.601	**0.794**
20,20	0.093	0.561	**0.910**	0.826	0.658	0.820
25,20	0.093	0.589	**0.905**	0.884	0.711	0.878
25,25	0.095	0.635	**0.972**	0.911	0.744	0.887
30,25	0.095	0.671	**0.971**	0.942	0.791	0.923
			$\theta = 0.4$			
15,10	0.098	0.455	0.656	0.697	0.571	**0.713**
15,15	0.110	0.516	**0.807**	0.744	0.617	**0.770**
20,15	0.116	0.581	**0.916**	0.852	0.702	**0.867**
20,20	0.124	0.656	**0.956**	0.890	0.750	0.886
25,20	0.126	0.691	**0.954**	0.937	0.805	0.933
25,25	0.139	0.747	**0.990**	0.957	0.846	0.943
30,25	0.143	0.779	**0.990**	0.977	0.877	0.967

TABLE 1.3: continued

			skewness=8			
m,n	t	W	KS	$log-tt$	$bc-ttL$	$bc-ttGoF$
			$\theta = 0$			
15,10	0.014	0.046	0.028	0.042	0.051	0.057
15,15	0.013	0.044	0.027	0.047	0.054	0.058
20,15	0.015	0.046	0.034	0.049	0.057	0.061
20,20	0.017	0.051	0.037	0.049	0.058	0.062
25,20	0.017	0.049	0.037	0.048	0.052	0.057
25,25	0.021	0.049	0.036	0.048	0.053	0.058
30,25	0.023	0.051	0.035	0.048	0.055	0.056
			$\theta = 0.1$			
15,10	0.026	0.668	**0.925**	0.915	0.852	0.761
15,15	0.026	0.769	**0.985**	0.940	0.909	0.949
20,15	0.024	0.814	**0.995**	0.982	0.946	0.955
20,20	0.025	0.882	**0.999**	0.990	0.969	0.983
25,20	0.021	0.904	**0.999**	0.998	0.985	0.991
25,25	0.026	0.941	**1.000**	0.998	0.992	0.996
30,25	0.025	0.949	**1.000**	**1.000**	0.996	0.998
			$\theta = 0.2$			
15,10	0.056	0.736	**0.956**	0.947	0.908	0.790
15,15	0.054	0.843	**0.995**	0.970	0.961	0.977
20,15	0.043	0.875	**0.999**	0.993	0.981	0.971
20,20	0.049	0.936	**1.000**	0.995	0.992	0.995
25,20	0.039	0.947	**1.000**	**1.000**	0.995	0.995
25,25	0.046	0.971	**1.000**	**1.000**	0.998	0.999
30,25	0.045	0.980	**1.000**	**1.000**	0.999	0.999

$\theta = 0.3$						
15,10	0.086	0.777	**0.972**	0.964	0.938	0.819
15,15	0.085	0.884	**0.996**	0.978	0.978	0.987
20,15	0.072	0.905	**1.000**	0.995	0.987	0.980
20,20	0.076	0.955	**1.000**	0.997	0.996	0.998
25,20	0.069	0.969	**1.000**	**1.000**	0.997	0.997
25,25	0.073	0.980	**1.000**	**1.000**	0.999	0.999
30,25	0.068	0.984	**1.000**	**1.000**	0.999	**1.000**
$\theta = 0.4$						
15,10	0.120	0.806	**0.979**	0.972	0.951	0.833
15,15	0.117	0.913	**0.998**	0.982	0.986	0.991
20,15	0.102	0.927	**1.000**	0.997	0.990	0.982
20,20	0.113	0.969	**1.000**	0.999	0.997	0.999
25,20	0.102	0.977	**1.000**	**1.000**	0.998	0.998
25,25	0.103	0.990	**1.000**	**1.000**	**1.000**	**1.000**
30,25	0.096	0.992	**1.000**	**1.000**	**1.000**	**1.000**

IV. EXAMPLE

Example 1

In this example, we generate to two samples of life expectancy of males (X) and females (Y) in the world to assess if there is any difference in the mean or median of male and female populations via parametric and non-parametric test within their applicability.

X: 47.0, 70.0, 39.4, 69.8, 52.3, 49.9, 72.3, 55.0, 50.3, 68.0, 67.5, 60.4, 57.5, 68.3, 73.3, 73.3, 67.2, 51.3, 43.4, 49.9, 71.8, 63.7, 70.7, 63.0, 51.0

Y: 72.8, 77.7, 52.2, 78.7, 66.1, 46.6, 80.0, 75.5, 52.1, 80.1, 74.5, 66.0, 69.2, 79.9, 66.5, 57.5, 51.0, 64.8, 71.6, 66.1, 75.4, 45.6, 74.9, 62.5, 72.7

Descriptive summary of the two datasets are as follows:

For sample X: $\overline{X} = 60.252$, $s_x = 10.47$, $Skewness = -0.34$, $Kurtosis = 1.77$, Shapiro.test of normality $P.value = 0.0326$

For sample Y: $\overline{Y} = 67.2$, $s_y = 10.80$, $Skewness = -0.66$, $Kurtosis = 2.25$, Shapiro.test of normality $P.value = 0.0229$

Tests*	Value of test statistic	P-value	Estimate of λ
t	-2.309	0.025	-
W	196.0	0.024	-
KS	0.360	0.078	-
$bc - ttL$	-2.445	0.018	2.3
$bc - ttGoF$	-2.535	0.015	3.2
$log - tt$	-2.202	0.032	-

* t =Untransformed T-test, W =Wilcoxon and Mann-Whitney test
KS = Kolmogorov-Smirnov test,
$bc - ttL$ =Box-Cox transformed t test with transformation parameter achieved via MLE method
$bc - ttGoF$ =Box-Cox transformed t test with transformation parameter achieved via Goodness-of-fit approach

Example 2

In this example, we generate to two samples of exam scores from two sections of a statistics class. We wish to investigate if there is any difference in the mean or median of scores of students in two classes with the various underlying tests.

X: 60, 95, 90, 50, 85, 100, 85, 85, 65, 90, 85, 95, 75, 95, 85, 85, 75, 70, 70, 80, 85, 80, 100, 90, 95, 50, 95, 95, 70, 85,

Y: 80, 85, 95, 95, 95, 90, 90, 95, 100, 95, 100, 80, 90, 95, 80, 95, 65, 90, 85, 74,

Descriptive summary of the two datasets are as follows:

For sample X: $\overline{X} = 82.2$, $s_x = 13.5$, $Skewness = -0.90$, $Kurtosis = 3.13$, Shapiro.test of normality $P.value = 0.0104$.

For sample Y: $\overline{Y} = 88.7$, $s_y = 9.1$, $Skewness = -1.0$, $Kurtosis = 3.43$, Shapiro.test of normality $P.value = 0.0236$

Tests	Value of test statistic	P-value	Estimate of λ
t	-1.894	0.064	-
W	214.50	0.087	-
KS	0.283	0.290	-
$bc - tt$	-1.826	0.074	3.9
$bc - ttGoF$	-1.822	0.075	4
$log - tt$	-1.886	0.065	-

.

V. RESULT DISCUSSION AND CONCLUSION

In application with examples 1, all tests seem to reject the null hypothesis at 5% level of significance, with the exception of the KS test, which seems to accept the null hypothesis. However, for example 2, all tests make identical conclusion with an acceptance of the null hypothesis at 5% level of significance.

Results of simulation from Tables 1.1-1.3, it follows that all tests demonstrate higher power as the mean difference θ increases. For simulations from beta distributions, the log-transformed t test have the best power among all tests; the highest power appears in bold for all tests. For simulations from exponential distributions, either the log-transformed t or $bc - ttGoF$ has the best power; often, for small samples $bc - ttGoF$ seems to perform better than log-transformed t, which gets reversed for moderate to large samples. For simulations from the gamma distributions with skewness=1, 2, the log-transformed t outperforms other tests in estimated power, whereas for skewness=4, 6, the KS test outperforms other tests in power, with some exceptions where $bc - ttGoF$ seems to perform better for small samples. All tests seem to have comparable estimated significance levels, except the untransformed t, which breaks down for higher skewness. Overall, we conclude that the log-transformed t test can be applied in any situation where NP or other transformed test can be used with acceptable success given the fact that it is the easiest transformed test to implement.

REFERENCES

[1]. G. Casella and R. Berger, Statistical Inference, second edition, Duxbury, 2002.
[2]. R. V. Hogg, J. W. Mckean and A. T. Craig, Introduction to Mathematical Statistics, Sixth edition, Pearson, Prentice Hall, 2005.
[3]. Ramsey, F. L. & Schafer, D. W. (2018). The Statistical Sleuth. A Course in Methods of Data Analysis, Third Edition, Cengage.
[4]. J. D. Gibbons. Nonparametric Statistical Inference, McGraw-Hill, 1971.
[5]. J. D. Gibbons and S. Chakraborti, Nonparametric Statistical Inference, Fourth Edition: Revised and Expanded. Taylor & Francis, 2014.
[6]. G. W. Corder and D. I. Foreman, Nonparametric Statistics: A Step-by-Step Approach, Second Edition, Wiley, 2014.
[7]. T. P. Hettmansperger and J. W. McKean, Robust Nonparametric Statistical Methods. Second Edition, CRC Press, 2010.
[8]. P. H. Kvam and B. Vidakovic, Nonparametric Statistics with Applications to Science and Engineering, Wiley, 2007.
[9]. P. Sprent and N. C. Smeeton, Applied Nonparametric Statistical Methods. Fourth Edition, CRC Press, 2014.
[10]. J. Kloke and J. W. McKean, Nonparametric Statistical Methods Using R, CRC Press, 2014.
[11]. M. Usman, On Consistency and Limitation of Independent t-test Kolmogorov Smirnov Test and Mann Whitney U test, IOSR Journal of Mathematics, vol.12, no.2, pp. 22-27, 2016.
[12]. G. E. P. Box and D. R. Cox, Analysis of transformation. Journal of Royal Statistical Society. Ser. B 26 211-246, 1964.
[13]. F. Mosteller and J. W. Tukey, Data Analysis and Regression, Addison-Wesley, Reading, MA, 1977.
[14]. P. J. Bickel and K. A. Doksum, An analysis of transformations revisited. J. Amer. Statist. Assoc. 76, 296-311, 1981.
[15]. A. C. Atkinson, Plots, Transformations and Regression. Oxford University Press, London, 1985.
[16]. K. Islam and T. Shapla, On Performance of Tests for Paired Samples: An Empirical Approach. IOSR
[17]. Journal of Mathematics, vol.14, no.2, pp. 36-47, 2018.
[18]. W. W. Daniel. Kolmogorov–Smirnov one-sample test. Applied Nonparametric Statistics, 2nd edition, Boston, 1990.

Khairul Islam. "Evaluation of Two-sample T-test and Analogous Nonparametric Tests: A Simulation Approach." *International Journal of Mathematics and Statistics Invention (IJMSI)*, vol. 09(01), 2021, pp. 01-11.