Two-sample Location Tests under Violation of the Normality and Variance Homogeneity Assumptions

Mongkol Leelaphaiboon [1] and Bumrungsak Phuenaree[2]*

[1]Department of Applied Mathematics and Statistics, Faculty of Sciences and Liberal Arts, Rajamangala University of Technology Isan, Nakhonratchasima, Thailand
[2]Department of Mathematics, Faculty of Science, Burapha University, Chonburi, Thailand

Received: 26 November 2020, Revised: 15 March 2021, Accepted: 8 April 2021

Abstract

In this research, the performance of four test statistics, the independent *t*-test, Welch's *t*-test, the Mann-Whitney test and the permutation test, were compared under combined violations of normality and homogeneity of variance. In a simulation study, we generated data from symmetric and asymmetric distributions. The results showed that all methods displayed reliable results in terms of protecting type I error rates at the nominal level, except for the Mann-Whitney test which provides an inflation of type I error rates. Considering the power of the tests for symmetric distributions with the homogeneity of variances, the independent *t*-test is the best test when the sample data are drawn from normal and uniform distributions, while the Mann-Whitney test is the most powerful for the logistic and Laplace distributions. With symmetric distributions in heterogeneity of variance cases, the permutation test is the most powerful test. For gamma distribution, the permutation test is the best test. In addition, this test is also the best option for the low degree of skewness for Log-normal distribution.

Keywords: permutation test; Welch's *t*-test; Mann-Whitney test; statistical power; type I error
DOI 10.14456/cast.2021.58

1. Introduction

Two sample *t*-test is one of the most frequently used approach in statistics. This method is a test of equality of two means. There are three conditions: a) normality assumption b) homogeneity of variances and c) independence of samples, that need to be examined before using this test. The independent *t*-test is derived under an equal variance situation. If two samples have an unequal variance, Welch's *t*-test is generally preferred.

Both independent *t*-test and Welch's *t*-test are robust tests when the first two assumptions were violated. However, there is no guarantee that *t*-test is the most powerful [1] and in this case, the other methods that non-parametric alternative approach should be performed. Mann-Whitney test is one of the most commonly used non-parametric statistical test for two samples. This test can

*Corresponding author: E-mail: bumrungsak@buu.ac.th

be used when the distributions are unknown; in other words, there is no normality assumption. Therefore, non-parametric tests are also called distribution free.

One of the non-parametric statistics that can be used to compute the sampling distribution for all test statistics is the permutation test. The permutation test does not need any assumptions. It gives a simple way to find the sampling distribution for all test statistics. If the null hypothesis is true, any observations from one group can be permuted to the other. The permutation test can be applied to many parametric statistics. In order to examine this test, the sampling distribution of the difference in means of two groups is considered in this work.

In some fields of research, especially in medical work and psychological data, the assumptions of normality and homogeneity of variance are often violated [1, 2]. Thus, the main purpose of this work is to compare the performance of four test statistics: the independent *t*-test, Welch's *t*-test, the Mann-Whitney test and the permutation test in order to figure out the best testing procedure. The non-normal data used in this study are symmetric and asymmetric distributions with varying degrees of standard deviation ratios.

2. Materials and Methods

This research studies four methods; the independent *t*-test, Welch's *t*-test, Mann-Whitney test and the permutation test, all of which can be used to compare location parameters in two populations. Consider two groups A and B. Let $X_1, X_2, ..., X_n$ be the observations of A, and $Y_1, Y_2, ..., Y_n$ be the observations of B. The details of each test are as follows.

2.1 Independent *t*-test

The independent two sample *t*-test is always used to compare two means when the population variances are equal. This test can be calculated as follows [2];

$$t = \frac{\bar{X}-\bar{Y}}{\sqrt{S_p^2\left(\frac{1}{n_X}+\frac{1}{n_Y}\right)}} \sim t_{n_X+n_Y-2} \quad (1)$$

$$S_p^2 = \frac{(n_X-1)S_X^2+(n_Y-1)S_Y^2}{n_X+n_Y-2}$$

where $\bar{X}$ and $\bar{Y}$ are the sample means, S_X^2 and S_Y^2 are the sample variances, and n_X and n_Y are the sample sizes.

2.2 Welch's *t*-test

The Welch's *t*-test is used to compare two means in the case of unequal variances [2]. This test is computed using the formula below:

$$W = \frac{\bar{X}-\bar{Y}}{\sqrt{\frac{S_X^2}{n_X}+\frac{S_Y^2}{n_Y}}} \sim t_\nu \quad (2)$$

where $\bar{X}$ and $\bar{Y}$ are the sample means, S_X^2 and S_Y^2 are the sample variances, and n_X and n_Y are the sample sizes.

The degree of freedom (ν) is given by [2]

$$\nu = \left(\frac{S_X^2}{n_X} + \frac{S_Y^2}{n_Y} \right)^2 \left(\frac{S_X^4}{n_X^3 - n_X^2} + \frac{S_Y^4}{n_Y^3 - n_Y^2} \right)^{-1} . \tag{3}$$

2.3 Mann-Whitney test

The Mann-Whitney test (MW) is a nonparametric test that is used when the two samples are not drawn from the normal distribution [3]. This test involves calculating

$$MW = \min(U_X, U_Y) \tag{4}$$

$$U_X = n_X n_y + n_X (n_X + 1) / 2 - R_X$$

$$U_Y = n_X n_y + n_Y (n_Y + 1) / 2 - R_Y$$

where n_X and n_Y are the sample sizes of the first and the second groups respectively, and R_X and R_Y are the sum of the ranks in samples X and Y.

When the observations are large enough, the statistic *MW* is approximately normal distributed with mean $n_X n_Y / 2$ and variance $n_X n_Y (n_X + n_Y + 1) / 12$. The test statistic becomes

$$z = \left(MW - \frac{n_X n_Y}{2} \right) \left(\sqrt{\frac{n_X n_Y (n_X + n_Y + 1)}{12}} \right)^{-1} . \tag{5}$$

2.4 Permutation test

Suppose that $X_1, X_2, ..., X_{n_x}$ and $Y_1, Y_2, ..., Y_{n_y}$ are $n_X + n_Y = N$ random samples from the first and the second groups, respectively. Considering N samples for this study, the n_X are randomly assigned to the first group, whereas the remaining $N - n_X$ will be assigned to the other group. There are $\binom{N}{n_X}$ possible randomizations. Then computed the difference in means, $D = \bar{X} - \bar{Y}$ for each of these randomizations [4]. The p-value can be calculated as

$$p - value = P\left(|D_i| \geq |D^*| \right) = \frac{\sum_{i=1}^{\binom{N}{n_X}} I(|D_i| \geq |D^*|)}{\binom{N}{n_X}} \tag{6}$$

where D_i difference in means for *i* th randomization and D^* is the difference in means of the observations. But if the samples are large, for example, if there are 10 observations in each sample, then over 184,000 randomizations are possible; $\binom{20}{10} = 184,756$. It is not easy to obtain all permutations in a short run-time computer program, so the p-value can be estimated with the Monte Carlo sampling from the permutation distribution [4]. The approximate *p*-value is

$$\hat{p} = \frac{1 + \sum_{i=1}^{B} I\left(\left|D_i\right| \geq \left|D^*\right|\right)}{B+1} \tag{7}$$

where B is permutation replications.

2.5 Simulation Study

This section provides simulation case studies for the type I error rates and the test powers of four statistics; the independent *t*-test (T), Welch's *t*-test (WT), the Mann-Whitney test (MW) and the permutation test (PER). The data were generated under six sampling distributions; normal, uniform, logistic, Laplace, gamma and lognormal distributions with balance sample sizes; n = 10, 15, 20, 25, 30, 50 and 100.

In order to examine the power of the test, two sets of the difference in parameters (Δ) were considered. The first set was {0, 1, 2}; location parameters for symmetric data, and the second set was {0, 0.25, 0.50, 0.75}; shape and scale parameters for skewed data. The effect of unequal variances for symmetric distributions were considered by defining the standard deviation ratios. These values were 1.0, 1.5, 2.0 and 2.5. The coefficients of skewness for gamma and lognormal distributions were 1 and 2. The summary of all distribution simulation cases are shown in Tables 1 and 2.

In this study, the Monte Carlo technique was performed using R version 3.4.1 [5]. The simulation and permutation trials were 10,000 and 2,000 respectively. The results for type I error rates and test powers are shown in Tables 3 -13.

Table 1. Summary of symmetric distribution simulation cases

Sampling distribution	Normal, Uniform, Logistic, Laplace
Difference in location parameters (means)	Δ = 0, 1, 2
Standard deviation ratios	1.0, 1.5, 2.0, 2.5
Method	T, WT, MW, PER
Equal sample sizes	10, 15, 20, 25, 30, 50, 100
Significance level	0.05

Table 2. Summary of skew distribution simulation cases

Sampling distribution	Gamma, Lognormal
Coefficient of skewness	1, 2
Difference in parameters	$\Delta = 0, 0.25, 0.50, 0.75$
Gamma (α, β) *	Group 1; Gamma (α, β)
$\beta = 1$	Group 2; Gamma $(\alpha, \beta + \Delta)$
Lognormal (μ, σ^2) **	Group 1; Lognormal (μ, σ^2)
$\mu = 1$	Group 2; Lognormal $(\mu + \Delta, \sigma^2)$
Method	WT, MW, PER
▪ Equal sample sizes	10, 15, 20, 25, 30, 50, 100
▪ Significance level	0.05

* Gamma (α, β) ; α and β are shape and scale parameters respectively.

** Lognormal (μ, σ^2) ; μ and σ^2 are location and shape parameters respectively.

***Both shape parameters; α and σ^2 , are defined as the coefficients of skewness.

3. Results and Discussion

For each studied situation, two criteria were used to examine the efficiency tests. The first criterion was the type I error rates $(\hat{\alpha})$, which should be close to the significance level of 0.05. The criterion of robustness was established on the Cochran's limit, that is $0.04 \le \hat{\alpha} \le 0.06$ for this work [6]. If the type I error rates are in this interval, it can be assumed that the rates are sufficiently close to the nominal level.

The second criterion was the power of the test. The methods that have the highest power are considered as the best among all the methods.

3.1 Type I error rates

In Tables 3, 4 and 7, it can be seen that type I error rates fell well within the range of Cochran's criteria. This implies that the rates for all test statistics are maintained near the nominal level regardless of the distribution shapes and sample sizes. In other words, they provide appropriate control of the type I error probability.

As seen in Tables 5 and 6, the type I error rates of the Mann-Whitney test increased when the variance ratio became larger. In other words, the Mann-Whitney test provides the inflation of type I error [7]. This type of results reveals the problem of this test. If the samples are selected randomly from two populations with the same means but with different variances, the type I error rates are far from the significance level in many cases. It shows the problem of lack of robustness of this test. In other words, the Mann-Whitney test is sensitive to population differences [8, 9]. Therefore, the Mann-Whitney test is not investigated in terms of the power values in these situations.

3.2 Power of the test

3.2.1 Symmetric Distribution in homogeneity of variance cases

Table 8 illustrates the power values of all tests for normal and uniform distributions. It can be clearly seen that all cases of the independent *t*-test have the highest power values. Moreover, the powers of all test are the same when the mean difference is 2 ($\Delta = 2$) and the sample sizes are greater than 15.

The details of the comparative study for logistic and Laplace distributions are shown in Table 9. The power of Mann-Whitney test is the highest when the mean difference is 1. However, all tests are powerful when the mean difference is 2 ($\Delta = 2$) and the sample sizes are greater than 20.

With the heterogeneity of variance in Tables 10 and 11, almost all cases of the permutation test have the highest power values. However, both tests are powerful when the sample sizes become large.

Table 3. Type I error rates for normal and logistic distributions in homogeneity of variance cases

	Normal			Logistic		
n	T	MW	PER	T	MW	PER
10	0.0491	0.0465	0.0514	0.0484	0.0430	0.0489
15	0.0527	0.0470	0.0518	0.0486	0.0469	0.0494
20	0.0485	0.0485	0.0483	0.0476	0.0462	0.0479
25	0.0519	0.0483	0.0508	0.0498	0.0482	0.0501
30	0.0495	0.0494	0.0494	0.0511	0.0491	0.0504
50	0.0466	0.0483	0.0467	0.0517	0.0484	0.0503
100	0.0535	0.0545	0.0534	0.0507	0.0479	0.0511

Table 4. Type I error rates for uniform and Laplace distributions in homogeneity of variance cases

	Uniform			Laplace		
n	T	MW	PER	T	MW	PER
10	0.0536	0.0430	0.0524	0.0461	0.0414	0.0499
15	0.0494	0.0424	0.0487	0.0525	0.0508	0.0540
20	0.0503	0.0486	0.0493	0.0499	0.0491	0.0504
25	0.0496	0.0495	0.0497	0.0510	0.0513	0.0528
30	0.0526	0.0524	0.0521	0.0528	0.0532	0.0553
50	0.0485	0.0488	0.0489	0.0476	0.0513	0.0480
100	0.0484	0.0486	0.0480	0.0492	0.0511	0.0488

Table 5. Type I error rates for normal and logistic distributions in heterogeneity of variance cases

$\frac{\sigma_2}{\sigma_1}$	n	Normal			Logistic		
		WT	MW	PER	WT	MW	PER
1.5	10	0.0497	0.0445	0.0520	0.0450	0.0454	0.0497
	15	0.0506	0.0516	0.0528	0.0494	0.0477	0.0518
	20	0.0496	0.0511	0.0516	0.0441	0.0513	0.0458
	25	0.0507	0.0535	0.0511	0.0515	0.0531	0.0521
	30	0.0512	0.0540	0.0521	0.0503	0.0530	0.0533
	50	0.0449	0.0449	0.0453	0.0507	0.0524	0.0513
	100	0.0534	0.0557	0.0532	0.0498	0.0533	0.0511
2.0	10	0.0457	0.0475	0.0501	0.0452	0.0466	0.0526
	15	0.0442	0.0484	0.0479	0.0439	0.0460	0.0503
	20	0.0464	0.0571	0.0490	0.0496	0.0598	0.0531
	25	0.0516	0.0585	0.0540	0.0500	0.0574	0.0538
	30	0.0493	0.0590	0.0508	0.0467	0.0582	0.0500
	50	0.0524	0.0585	0.0529	0.0473	0.0578	0.0490
	100	0.0471	0.0577	0.0473	0.0525	0.0601	0.0517
2.5	10	0.0483	0.0568	0.0562	0.0555	0.0561	0.0576
	15	0.0515	0.0581	0.0560	0.0482	0.0567	0.0569
	20	0.0482	0.0611	0.0527	0.0486	0.0498	0.0510
	25	0.0510	0.0638	0.0543	0.0501	0.0664	0.0545
	30	0.0480	0.0620	0.0504	0.0486	0.0598	0.0517
	50	0.0510	0.0633	0.0534	0.0486	0.0616	0.0503
	100	0.0493	0.0627	0.0497	0.0513	0.0638	0.0518

Table 6. Type I error rates for uniform and Laplace distributions in heterogeneity of variance cases

$\frac{\sigma_2}{\sigma_1}$	n	Uniform			Laplace		
		WT	MW	PER	WT	MW	PER
1.5	10	0.0514	0.0491	0.0517	0.0414	0.0418	0.0486
	15	0.0502	0.0516	0.0511	0.0480	0.0477	0.0526
	20	0.0504	0.0562	0.0506	0.0478	0.0514	0.0526
	25	0.0561	0.0614	0.0560	0.0471	0.0515	0.0499
	30	0.0498	0.0559	0.0502	0.0488	0.0509	0.0501
	50	0.0529	0.0595	0.0532	0.0488	0.0528	0.0504
	100	0.0494	0.0558	0.0499	0.0457	0.0470	0.0455
2.0	10	0.0540	0.0566	0.0557	0.0401	0.0466	0.0515
	15	0.0540	0.0630	0.0580	0.0449	0.0502	0.0517
	20	0.0563	0.0633	0.0555	0.0481	0.0542	0.0525
	25	0.0453	0.0586	0.0468	0.0488	0.0572	0.0533
	30	0.0518	0.0647	0.0526	0.0488	0.0515	0.0514
	50	0.0491	0.0644	0.0491	0.0452	0.0546	0.0483
	100	0.0458	0.0615	0.0464	0.0487	0.0548	0.0503
2.5	10	0.0554	0.0610	0.0597	0.0469	0.0484	0.0528
	15	0.0535	0.0672	0.0578	0.0457	0.0510	0.0560
	20	0.0517	0.0682	0.0547	0.0482	0.0580	0.0545
	25	0.0496	0.0699	0.0525	0.0490	0.0572	0.0541
	30	0.0502	0.0707	0.0533	0.0452	0.0563	0.0509
	50	0.0529	0.0700	0.0543	0.0472	0.0615	0.0496
	100	0.0500	0.0706	0.0500	0.0513	0.0589	0.0526

Table 7. Type I error rates for skewed distribution

		Gamma			Lognormal		
Skewness	n	WT	MW	PER	WT	MW	PER
1	10	0.0445	0.0427	0.0510	0.0492	0.0461	0.054
	15	0.0496	0.0466	0.0533	0.0471	0.0441	0.0493
	20	0.0480	0.0471	0.0500	0.0531	0.0527	0.0541
	25	0.0468	0.0478	0.0478	0.0522	0.0489	0.0523
	30	0.0506	0.0525	0.0518	0.0510	0.0512	0.0519
	50	0.0483	0.0483	0.0494	0.0518	0.0509	0.0520
	100	0.0536	0.0513	0.0544	0.0501	0.0510	0.0493
2	10	0.0405	0.0459	0.0497	0.0407	0.0434	0.0509
	15	0.0426	0.0445	0.0501	0.0439	0.0447	0.0504
	20	0.0451	0.0517	0.0505	0.0461	0.0497	0.0520
	25	0.0478	0.0509	0.0507	0.0452	0.0481	0.0483
	30	0.0465	0.0462	0.0502	0.0464	0.0489	0.0494
	50	0.0465	0.0501	0.0486	0.0513	0.0502	0.0524
	100	0.0490	0.0516	0.0497	0.0478	0.0503	0.0487

Table 8. Power values for normal and uniform distributions in homogeneity of variance cases

n	$\Delta = 1$					
	Normal			Uniform		
	T	MW	PER	T	MW	PER
10	0.5645*	0.5126	0.5633	0.5381*	0.4632	0.5365
15	0.7553*	0.7171	0.7543	0.7502*	0.6675	0.7487
20	0.8646*	0.8440	0.8639	0.8721*	0.8080	0.8711
25	0.9334*	0.9211	0.9327	0.9397*	0.8869	0.9379
30	0.9686*	0.9602	0.9684	0.9714*	0.9402	0.9701
50	0.9979*	0.9973	0.9979*	0.9991*	0.9947	0.9989
100	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*
n	$\Delta = 2$					
	Normal			Uniform		
	T	MW	PER	T	MW	PER
10	0.9884*	0.9805	0.9881	0.9933*	0.9712	0.9930
15	0.9992*	0.9991	0.9992*	1.0000*	0.9975	1.0000*
20	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*
25	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*
30	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*
50	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*
100	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*

* Tests with the highest power value

Table 9. Power values for logistic and Laplace distributions in homogeneity of variance cases

n	$\Delta = 1$					
	Logistic			Laplace		
	T	MW	PER	T	MW	PER
10	0.5722	0.5545*	0.5743	0.5892	0.6216*	0.5942
15	0.7542	0.7572*	0.7537	0.7588	0.8213*	0.7609
20	0.8672	0.8837*	0.8670	0.8661	0.9283*	0.8661
25	0.9319	0.9453*	0.9320	0.9272	0.9714*	0.9286
30	0.9699	0.9774*	0.9702	0.9605	0.9900*	0.9610
50	0.9977	0.9992*	0.9977	0.9979	0.9997*	0.9977
100	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*
n	$\Delta = 2$					
	Logistic			Laplace		
	T	MW	PER	T	MW	PER
10	0.9834*	0.9817	0.9834*	0.9751	0.9785*	0.9752
15	0.9988	0.9995*	0.9988	0.9982	0.9987*	0.9983
20	1.0000*	0.9999	1.0000*	0.9996	1.0000*	0.9996
25	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*
30	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*
50	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*
100	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*

* Tests with the highest power value

Table 10. Power values for symmetric distribution in heterogeneity of variance cases with $\Delta = 1$

$\frac{\sigma_2}{\sigma_1}$	n	Normal		Logistic		Uniform		Laplace	
		WT	PER	WT	PER	WT	PER	WT	PER
1.5	10	0.3831	0.3949*	0.3889	0.4036*	0.3503	0.3541*	0.4177	0.4385*
	15	0.5426	0.5475*	0.5440	0.5518*	0.5171	0.5187*	0.5696	0.5815*
	20	0.6727	0.6754*	0.6733	0.6777*	0.6644	0.6655*	0.6872	0.6961*
	25	0.7729	0.7742*	0.7717	0.7742*	0.7680	0.7684*	0.7757	0.7833*
	30	0.8414	0.8430*	0.8496	0.8511*	0.8477	0.8490*	0.8395	0.8417*
	50	0.9720	0.9722*	0.9687	0.9692*	0.9762	0.9765*	0.9712	0.9713*
	100	0.9996*	0.9996*	0.9998*	0.9997	0.9999*	0.9999*	0.9999*	0.9999*
2.0	10	0.2615	0.2783*	0.2721	0.2933*	0.2336	0.2447*	0.2968	0.3269*
	15	0.3700	0.3857*	0.3939	0.4068*	0.3616	0.3716*	0.4234	0.4400*
	20	0.4824	0.4940*	0.5044	0.5156*	0.4812	0.4902*	0.5248	0.5378*
	25	0.5887	0.5963*	0.5921	0.5995*	0.5724	0.5792*	0.6042	0.6144*
	30	0.6657	0.6704*	0.6771	0.6840*	0.6531	0.6556*	0.6803	0.6898*
	50	0.8711	0.8727*	0.8807	0.8827*	0.8824	0.8848*	0.8761	0.8800*
	100	0.9934*	0.9934*	0.9937	0.9940*	0.9940	0.9943*	0.9913	0.9918*

Table 10. (cont.)

$\frac{\sigma_2}{\sigma_1}$	n	Normal		Logistic		Uniform		Laplace	
		WT	PER	WT	PER	WT	PER	WT	PER
2.5	10	0.1900	0.2118*	0.1128	0.1399*	0.1747	0.1918*	0.2257	0.2536*
	15	0.2790	0.2966*	0.1515	0.1729*	0.2581	0.2705*	0.3059	0.3277*
	20	0.3636	0.3784*	0.1853	0.2006*	0.3490	0.3596*	0.3883	0.4067*
	25	0.4403	0.4507*	0.2232	0.2378*	0.4201	0.4303*	0.4711	0.4884*
	30	0.5095	0.5197*	0.2597	0.2711*	0.4977	0.5039*	0.5245	0.5365*
	50	0.7389	0.7405*	0.7395	0.7445*	0.7364	0.7411*	0.7425	0.7476*
	100	0.9583	0.9588*	0.9587	0.9588*	0.9576	0.9587*	0.9544	0.9549*

* Tests with the highest power value

Table 11. Power values for symmetric distribution in heterogeneity of variance cases with $\Delta = 2$

$\frac{\sigma_2}{\sigma_1}$	n	Normal		Logistic		Uniform		Laplace	
		WT	PER	WT	PER	WT	PER	WT	PER
1.5	10	0.9056	0.9119*	0.9040	0.9115*	0.9186	0.9207*	0.8876	0.8954*
	15	0.9843	0.9852*	0.9760	0.9770*	0.9896	0.9895*	0.9753	0.9770*
	20	0.9977	0.9980*	0.9966	0.9969*	0.9991	0.9992*	0.9932	0.9934*
	25	0.9999*	0.9999*	0.9994*	0.9992	0.9999*	0.9999*	0.9989	0.9990*
	30	1.0000*	1.0000*	0.9999*	0.9999*	1.0000*	1.0000*	0.9997	0.9998*
	50	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*
	100	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*
2.0	10	0.7377	0.7577*	0.7528	0.7736*	0.7366	0.7530*	0.7615	0.7855*
	15	0.9120	0.9180*	0.9032	0.9118*	0.9201	0.9240*	0.8949	0.9020*
	20	0.9734	0.9748*	0.9670	0.9697*	0.9763	0.9774*	0.9573	0.9615*
	25	0.9906	0.9911*	0.9878	0.9889*	0.9947*	0.9947*	0.9844	0.9862*
	30	0.9978	0.9979*	0.9964	0.9967*	0.9987*	0.9987*	0.9963	0.9967*
	50	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	0.9999*	0.9999*
	100	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*
2.5	10	0.5711	0.6054*	0.3026	0.3477*	0.5534	0.5802*	0.6183	0.6566*
	15	0.7789	0.7966*	0.4462	0.4780*	0.7695	0.7816*	0.7911	0.8090*
	20	0.8929	0.9001*	0.5546	0.5787*	0.8921	0.8973*	0.8849	0.8939*
	25	0.9454	0.9483*	0.6573	0.6739*	0.9522	0.9540*	0.9378	0.9432*
	30	0.9781	0.9790*	0.7409	0.7527*	0.9790	0.9797*	0.9699	0.9716*
	50	0.9990	0.9991*	0.9989*	0.9988	0.9997*	0.9997*	0.9989*	0.9989*
	100	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*

* Tests with the highest power value

3.2.2 Skew distribution

Considering the skew distribution with coefficient of skewness varying; 1 and 2, the permutation test behaves better than the other two when both samples come from gamma distributions (Table 12).

Table 13 shows the power values of all tests for lognormal data. The permutation test gives the highest power when the coefficients of skewness are 1. However, the test power of all test statistics reaches to 1 when the sample sizes are greater than 10 with the high difference in location parameters. For high degree of skewness, the permutation test gives the best results when the sample sizes are 10. But, the Mann-Whitney test becomes the best test when the sample sizes are at least 15.

Table 12. Power values for Gamma distribution

			Skewness =1			Skewness =2	
Δ	n	WT	MW	PER	WT	MW	PER
0.25	10	0.1352	0.1279	0.1484*	0.0563	0.0636	0.0731*
	15	0.2012	0.1833	0.2108*	0.0764	0.0753	0.0882*
	20	0.2650	0.2463	0.2721*	0.0931	0.0863	0.1012*
	25	0.3345	0.3103	0.3348*	0.1095	0.1020	0.1173*
	30	0.3905	0.3627	0.3954*	0.1198	0.1131	0.1281*
	50	0.5999	0.5591	0.6013*	0.1859	0.1601	0.1907*
	100	0.8853*	0.8448	0.8853*	0.3426	0.2768	0.3434*
0.5	10	0.3679	0.3365	0.3922*	0.0964	0.1020	0.1288*
	15	0.5494	0.5056	0.5634*	0.1483	0.1418	0.1750*
	20	0.6874	0.6449	0.6954*	0.2040	0.1867	0.2239*
	25	0.7927	0.7471	0.7957*	0.2582	0.2249	0.2724*
	30	0.8630	0.8287	0.8644*	0.3147	0.2698	0.3268*
	50	0.9778	0.9645	0.9780*	0.4972	0.4125	0.5042*
	100	1.0000*	0.9999	1.0000*	0.8085	0.6867	0.8103*
0.75	10	0.6099	0.5798	0.6459*	0.1479	0.1537	0.1970*
	15	0.8103	0.7698	0.8205*	0.2525	0.2340	0.2879*
	20	0.9242	0.9012	0.9277*	0.3554	0.3109	0.3841*
	25	0.9686	0.9520	0.9692*	0.4469	0.3861	0.4694*
	30	0.9884	0.9803	0.9887*	0.5368	0.4522	0.5530*
	50	1.0000*	0.9998	1.0000*	0.7659	0.6560	0.7719*
	100	1.0000*	1.0000*	1.0000*	0.9733	0.9179	0.9737*

* Tests with the highest power value

Table 13. Power values for Log-normal distribution

		Skewness =1			Skewness =2		
Δ	n	WT	MW	PER	WT	MW	PER
0.25	10	0.3694	0.3575	0.3872*	0.1275	0.1365	0.1541*
	15	0.5300	0.5168	0.5392*	0.1933	0.2048	0.2123*
	20	0.6664	0.6606	0.6725*	0.2574	0.2821*	0.2720
	25	0.7682	0.7601	0.7696*	0.3012	0.3262*	0.3109
	30	0.8361	0.8377*	0.8369	0.3611	0.3948*	0.3684
	50	0.9710*	0.9696	0.9709	0.5577	0.5911*	0.5620
	100	0.9997*	0.9996	0.9997*	0.8431	0.8801	0.8440
0.5	10	0.9012	0.8915	0.9104*	0.4110	0.4406	0.4642*
	15	0.9825	0.9800	0.9834*	0.6009	0.6309*	0.6307
	20	0.9973	0.9968	0.9977*	0.7279	0.7714*	0.7451
	25	0.9996*	0.9996*	0.9996*	0.8244	0.8612*	0.8337
	30	1.0000*	1.0000*	1.0000*	0.8917	0.9198*	0.8918
	50	1.0000*	1.0000*	1.0000*	0.9874	0.9932*	0.9880
	100	1.0000*	1.0000*	1.0000*	0.9998	1.0000*	0.9998
0.75	10	0.9980	0.9979	0.9984*	0.7345	0.7825	0.7922*
	15	1.0000*	1.0000*	1.0000*	0.9076	0.9248*	0.9246
	20	1.0000*	1.0000*	1.0000*	0.9698	0.9805*	0.9743
	25	1.0000*	1.0000*	1.0000*	0.9908	0.9963*	0.9918
	30	1.0000*	1.0000*	1.0000*	0.9979	0.9993*	0.9983
	50	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*
	100	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*

* Tests with the highest power value

4. Conclusions

Based on the numerical studies from the previous section, increasing sample size is found in this study to improve the test power for all testing procedures, but the standard deviation ratios seem to have the different kinds of impact. In other words, the power values drop when the standard deviation ratios increase. Moreover, the power values of low skewness are greater than those of high skewness.

The results for the homogeneity of variance demonstrate that the independent *t*-test is a better test than the other two when the sample data are drawn from the normal and uniform distributions, while the Mann-Whitney test is the most powerful for the logistic and Laplace distributions. However, all tests perform well when the mean differences and the sample sizes become large.

With the symmetric distribution in heterogeneity of variance cases, the permutation test is more powerful than the Welch *t*-test. Moreover, both tests reach the same power values when the sample sizes become large. However, the Mann-Whitney test is not appropriate because the concept of this test is to test that two samples drawn from the same distribution; same means and same variances.

Instead of considering the difference in means, we consider the difference in parameters for skew distribution; the scale and shape parameters. So, the Mann-Whitney test can be examined in this case. For the gamma distribution, the permutation test is the best test. In addition, this test is also the best option in the case of low degree of skewness for log-normal distribution.

In conclusion, the concepts of all test are different. The Welch's *t*-test and the permutation test should be used to compare the central tendency of two populations, whereas the Mann-Whitney test should always be used to investigate two populations that are identical distribution. Of course, researchers should adopt the procedure that corresponds best with the objectives of their research design.

References

[1] Boneau, C.A., 1960. The effects of violations of assumptions underlying the t test. *Psychological Bulletin*, 57(1), 49-64.

[2] Fagerland, M.W. and Sandvik L., 2009. Performance of five two-sample location tests for skewed distributions with unequal variances. *Contemporary Clinical Trials*, 30, 490-496.

[3] Nadim, N., 2008. The Mann-Whitney U: a test for assessing whether two independent samples come from the same distribution. *Tutorials in Quantitative Methods for Psychology*, 4(1), 13-20.

[4] Ernst, M.D., 2004. Permutation methods: a basis for exact inference. *Statistical Science*, 19(4), 676-685.

[5] R Development Core Team, 2019. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

[6] Hatchavanich, D., 2014. A comparison of type I error and power of Bartlatt's test, Levene's test and O'Brien's test for homogeneity of variance test. *Southeast-Asian Journal of Sciences*, 3(2), 181-194.

[7] Kasuya, E., 2001. Mann-Whitney *U* test when variances are unequal. *Animal Behaviour*, 61, 1247-1249.

[8] Stonehouse, J.M. and Forrester, G.J., 1998. Robustness of the t and U tests under combined assumption violations. *Journal of Applied Statistics*, 25(1), 63-74.

[9] Fagerland, M.W., 2012. t-tests, non-parametric tests, and large studies-a paradox of statistical practice? *BMC Medical Research Methodology*, 12, 78, https://doi.org/10.1186/1471-2288-12-78.