Comparison of Test Statistics for Mean Difference Testing Between Two Independent Populations

Wasinee Pradubsri[1], Chawanee Suphirat[2,*]

[1]*Department of Science and Mathematics, Rajamangala University of Technology Isan Surin Campus, Surin 32000, Thailand*

[2]*Department of Mathematics and Statistics, Faculty of Science and Technology, Rajamangala University of Technology Phra Nakhon, Bangkok 10800, Thailand*

Corresponding author: chawanee.s@rmutp.ac.th

ABSTRACT. The purpose of the article is to evaluate the efficiency of seven test statistics for mean difference testing between two independent populations. The evaluation was based on the probability of type I error and power of the test at 0.05 significance level under population distributions assumed to be normal, exponential, log-normal, gamma, and Laplace with equal sample sizes, and both equal and unequal variances. The results showed that for equal variance, the test statistics with the highest testing power controlled the probability of type I error were Z-test for normal and exponential distributions, Welch based on rank test (WBR) for log-normal and gamma distributions, and Mann-Whitney U test (MWU) for Laplace distribution. For unequal variance, Z-test was more efficient under normal, exponential, log-normal, and gamma distributions, while WBR was appropriate for Laplace distribution.

1. Introduction

The testing of mean differences obtained from two independent populations have been extensively utilized for testing statistical hypotheses in a variety of research areas, especially in educational science and medical studies. Generally, there are various test statistics; therefore, the selection of appropriate test statistics corresponding to the properties of data sets and the objectives of the research are important for mean comparison tests. If the distributions of two populations from which samples are drawn are normal and known variances, Z-test is used. However, if the distributions of two populations have normal and unknown variances, t-test is

applied in case of homogeneity of variances, Welch t-test is used in case of heterogeneity of variances. These test statistics are called parametric statistics, which strongly depend on assumption of normality [5]. In practice, the populations that are non-normally distributed lead to possible erroneous conclusions. Consequently, non-parametric statistics have become an interesting alternative method when the normality assumption is violated ([8],[13]).

In literature, comparison of performance parametric and non-parametric statistics for testing means between two independent populations have been discussed by numerous publications. Thidarat et al. [12] compared the efficiency of test statistics of Mann-Whitney U test (MWU), Brunner and Munzel test, and nonparametric bootstrap rank Welch test (BRW) in terms of the probability of type I error and power of the test when the two populations were assumed to be normal, gamma, exponential, and Chi-squared distributions. They recommended that BRW had the highest testing power but the lowest ability to control the probability of type I error, whereas MWU was the second most powerful and able to control the probability of type I error if the variances were assumed to be equal. Eriobu and Umeh [4] pointed that MWU was better than Kolmogorov-Smirnov test and modified intrinsically ties adjusted Mann-Whitney U test (MAMWU) for the testing of two population means when populations had gamma and Weibull distributions. Sangthong and Klubnual [11] found that Welch based on rank test (WBR) was appropriate when the populations had log-normal, gamma, and poisson distributions with equal variance, while Welch t test yielded the highest efficiency when the populations had log-normal, gamma, exponential, uniform and poisson distributions with unequal variance. Dollada et al. [2] stated that MWU was the most powerful when the populations had negatively skewed distribution, symmetrical and leptokurtic distribution, and positively skewed and leptokurtic distribution.

In this article, the researchers aim to evaluate the performance of test statistics for mean differences between two independent populations. The test statistics compared are Z-test, t-test, Welch t-test, MWU, MAMWU, WBR, and BRW. A simulation study is implemented to verify that their test statistics have the ability to control the probability of type I error and the highest testing power under normal, exponential, log-normal, gamma, and Laplace distributions with equal and unequal variances at 0.05 significance level.

The rest of this article is organized as follows: Section 2 introduces the test statistics of this study. Then the construction method of these test statistics is provided in Section 3. To assess the efficiency of the test statistics, Section 4 presents simulated results to investigate the probability of type I error and power of the test. Finally, Section 5 presents the conclusion and discussion.

2. Statistical Methods

2.1 Z-test

Z -test is a parametric statistic used for testing the difference between means of two independent populations from a normal distribution with known variance [6]. The Z-test is:

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} \tag{2.1}$$

where μ_1, μ_2 are the means of populations 1 and 2. σ_1, σ_2 are the standard deviations of populations 1 and 2. n_1, n_2 are the sample sizes drawn from populations 1 and 2 and $\bar{x}_1$, $\bar{x}_2$ are the means of the samples drawn from populations 1 and 2.

Then reject H_0; when $Z \leq -Z_{\alpha/2}$ or $Z \geq Z_{\alpha/2}$.

2.2 t-test

The t-test for two populations, also known as the independent samples t-test, is a parametric statistic that is appropriate when you want to determine whether there is a significant difference between the means of the two populations if the population standard deviations are unknown and the samples are relatively small. There are two variations of the t-test for two populations: t-test for equal variances and Welch t-test for unequal variances.

The independent sample t-test is applied under the assumption of normal distribution and equal population variances [7]. The t-test statistic is:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{s_p^2 \left(\dfrac{1}{n_1} + \dfrac{1}{n_2} \right)}} \tag{2.2}$$

where $s_p^2 = \dfrac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$, and s_1, s_2 are the standard deviation of the sample drawn from populations 1 and 2.

Then reject H_0; when $t \leq -t_{\alpha/2,df}$ or $t \geq t_{\alpha/2,df}$ by $df = n_1 + n_2 - 2$.

2.3 Welch t-test

The Welch t-test was presented by Welch [14]. This test statistic is used to determine the significant of mean difference between two independent populations under the assumption of normal distribution and unequal population variances. The Welch t-test statistic is:

$$Welch\ t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}} \tag{2.3}$$

where $s_1^2 = \dfrac{\sum\limits_{i=1}^{n_1}(x_{i1} - \bar{x}_1)^2}{n_1 - 1}$ and $s_2^2 = \dfrac{\sum\limits_{i=1}^{n_2}(x_{i2} - \bar{x}_2)^2}{n_2 - 1}$.

Then reject H_0; when $Welch \ t \le -t_{\alpha/2,df}$ or $Welch \ t \ge t_{\alpha/2,df}$ by $df = \dfrac{\left(\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}\right)^2}{\dfrac{\left(s_1^2/n_1\right)^2}{n_1 - 1} + \dfrac{\left(s_2^2/n_2\right)^2}{n_2 - 1}}$.

2.4 Mann-Whitney U test (MWU)

Mann-Whitney U test was proposed by Wilcoxon [15] based on rank statistics assigned to observations from two populations. This test statistic is useful for testing hypothesis by assigning the same distribution from two independent populations [4].

For $n_1 \le 20$ and $n_2 \le 20$, the MWU test statistic is:

$$U = min(U_1, U_2) \tag{2.4}$$

where $U_1 = n_1 n_2 + \dfrac{n_1(n_1 + 1)}{2} - \sum R_1$ and $U_2 = n_1 n_2 + \dfrac{n_2(n_2 + 1)}{2} - \sum R_2$.

n_1, n_2 are the sample sizes drawn from populations 1 and 2 and $\sum R_1$, $\sum R_2$ are the sum of ranks assigned to observations from populations 1 and 2 in the combined ranking of these observations from the two populations.

Then reject H_0; when $U \le U_{\alpha/2}$

For $n_1 > 20$ and $n_2 > 20$, the MWU test statistic is:

$$Z = \dfrac{U - \mu_U}{\sigma_U} \tag{2.5}$$

where $U = n_1 n_2 + \dfrac{n_1(n_1 + 1)}{2} - \sum R_1$, $\mu_U = \dfrac{n_1 n_2}{2}$, and $\sigma_U = \sqrt{\dfrac{n_1 n_2(n_1 + n_2 + 1)}{12}}$

Then reject H_0; when $Z \le -Z_{\alpha/2}$ or $Z \ge Z_{\alpha/2}$.

2.5 Modified intrinsically ties adjusted Mann-Whitney U test (MAMWU)

Oyeka and Okeh [9] developed modified intrinsically ties adjusted Mann-Whitney U test (MAMWU), which is adjusted from the approach of MWU to solve the problem of tied observations between the two-sample populations. MAMWU is useful for hypothesis testing in which the sample sizes are drawn from the same distribution [4]. The MAMWU test statistic is:

$$\lambda^2 = \dfrac{(n_2 R_{.1} - n_1 R_{.2})^2}{(n_2 R_1^{*2} + n_1 R_2^{*2} - 2R_{.1} R_{.2})(\pi^+ + \pi^- - (\pi^+ + \pi^-)^2)} \tag{2.6}$$

where n_1, n_2 are the sample sizes drawn from populations 1 and 2. R_1, R_2 are the respective sums of the ranks assigned to observations from populations 1 and 2 in the combined ranking of these observations from the two populations, and π^+, π^- are, respectively, the probabilities that observations or scores by subject from population 1 is on the average greater than or less than observations or scores by subject from population 2.

Then reject H_0; when $\lambda^2 \geq \chi^2_{1-\alpha,1}$

2.6 Welch based on rank test (WBR)

Welch [14] proposed the Welch based on rank, which is associated with the midrank. It was explained as follows ([10], [11], [16]):

(1) Combine n_1 and n_2 and rank the data from the lowest to the highest. If tied observations between two samples are found, the midrank are averaged.

(2) Calculate the means and variances of midrank to the first and second samples, denoted as $\bar{R}_1$ and $\bar{R}_2$, and calculated s_1^2 and s_2^2, respectively.

(3) Calculate the Welch based on rank test statistic (WBR):

$$WBR = \frac{\bar{R}_1 - \bar{R}_2}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}} \tag{2.7}$$

(4) Reject H_0; when $WBR \leq -t_{\alpha/2,df}$ or $WBR \geq t_{\alpha/2,df}$ by $df = \dfrac{\left(\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}\right)^2}{\dfrac{\left(s_1^2/n_1\right)^2}{n_1-1} + \dfrac{\left(s_2^2/n_2\right)^2}{n_2-1}}$.

2.7 Nonparametric bootstrap rank Welch test (BRW)

The nonparametric bootstrap rank Welch test (BRW) is applied from the principle of bootstrap and the technique of rank Welch test statistic ([3], [10]). The method of hypothesis testing is the following:

(1) Let $x_1 = x_{11}, x_{12}, x_{13}, ..., x_{1n_1}$ be the observations in sample 1 of size n₁ drawn from population 1 and $x_2 = x_{21}, x_{22}, x_{23}, ..., x_{2n_2}$ be the observations in sample 2 of size n₂ drawn from population 2.

(2) Pool the two samples into one combined sample of size $n = n_1 + n_2$, then ascend order. If tied observations between two samples are found, the midrank are averaged.

(3) Calculate the WBR test statistics in equation (2.7).

(4) Return sampling with replacement from x_1 and x_2 of size n_1 and n_2 transform into x_1^* and x_2^*, then ascend order for the combined observations. If tied observations between two samples are found, the midrank are averaged.

(5) Calculate the WBR* test statistics:

$$WBR^* = \frac{\bar{R}_1^* - \bar{R}_2^*}{\sqrt{\dfrac{s_1^{2^*}}{n_1} + \dfrac{s_2^{2^*}}{n_2}}} \tag{2.8}$$

(6) Repeat step (4) and (5) for B iterations (B = 10,000).

(7) Approximate p value $= \dfrac{\text{number of times}\left(\left|WBR^*\right| \geq \left|WBR\right|\right)}{B}$

3. Research Method

The Monte Carlo simulation techniques were used to generate data for determining the performance of test statistics. MATLAB version R2021b was written to calculate the probability of type I error and power of the test for the following conditions:

1) Generate two independent populations into five distributions:

Distribution	The probability of type I error calculation	Power of the test calculation
normal	$X_1 \sim N(\mu_1 = 1, \sigma_1^2 = 1)$ $X_2 \sim N(\mu_2 = 1, \sigma_2^2 = a)$	$X_1 \sim N(\mu_1 = 1, \sigma_1^2 = 1)$ $X_2 \sim N(\mu_2 = M, \sigma_2^2 = a)$
exponential	Equal Variance: $\sigma^2 = 1$ $X_1 \sim Exp(\mu_1 = 1)$ $X_2 \sim \sigma\, Exp(\mu_2 = 1)$ Unequal Variance: $\sigma^2 = 3, 5, 9$ $X_1 \sim Exp(\mu_1 = 1) - 1$ $X_2 \sim \sigma\left[Exp(\mu_2 = 1) - 1\right]$	$X_1 \sim Exp(\mu_1 = 1)$ $X_2 \sim \sigma\left[Exp(\mu_2 = 1) - 1 + M\right]$ $\sigma^2 = 1, 3, 5, 9$.
log-normal	Equal Variance: $\sigma^2 = 1$ $X_1 \sim Lognorm(\mu_1 = 1, \sigma_1^2 = 1) - exp^{3/2} + 1$ $X_2 \sim \sigma\left[Lognorm(\mu_2 = 1, \sigma_2^2 = 1) - exp^{3/2} + 1\right]$ Unequal Variance: $\sigma^2 = 3, 5, 9$ $X_1 \sim Lognorm(\mu_1 = 1, \sigma_1^2 = 1) - exp^{3/2}$ $X_2 \sim \sigma\left[Lognorm(\mu_2 = 1, \sigma_2^2 = 1) - exp^{3/2}\right]$	$X_1 \sim Lognorm(\mu_1 = 1, \sigma_1^2 = 1) - exp^{3/2} + 1$ $X_2 \sim \sigma\left[Lognorm(\mu_2 = 1, \sigma_2^2 = 1) - exp^{3/2} + M\right]$ $\sigma^2 = 1, 3, 5, 9$.

gamma	$X_1 \sim gamma(\alpha_1 = a, \beta_1 = 1)$ $X_2 \sim gamma(\alpha_2 = 1, \beta_2 = a)$	$X_1 \sim gamma(\alpha_1 = 1, \beta_1 = 1) - 1 + 1$ $X_2 \sim \sigma[gamma(\alpha_2 = 1, \beta_2 = a) - 1 + M]$ $\sigma^2 = 1, 3, 5, 9$.
Laplace	$X_1 \sim Laplace(\mu_1 = 1, \sigma_1^2 = 1)$ $X_2 \sim \sigma\, Laplace(\mu_1 = 1, \sigma_2^2 = 1)$ $\sigma^2 = 1, 3, 5, 9$.	$X_1 \sim Laplace(\mu_1 = 1, \sigma_1^2 = 1) + 1$ $X_2 \sim \sigma\left[Laplace(\mu_2 = 1, \sigma_2^2 = 1) - 1 + M\right]$ $\sigma^2 = 1, 3, 5, 9$.

Where $a = 1, 3, 5, 9$ and $M = 1.5, 2$.

2) Determine the equal sample sizes for two populations. $(n_1, n_2) = (10, 10)$, (20,20), and (50,50).

3) Assign the difference of means for two population groups as follows:

3.1) The means between two population groups are not different when calculating the probability of type I error.

3.2) The means between two population groups are different when computing power estimation with the ratios of 1:1.5 and 1:2 [13].

4) Specify four variance ratios for both equal and unequal variances. The variance ratios were 1:1, 1:3, 1:5, and 1:9.

5) The significance level of the test is 0.05 ($\alpha = 0.05$).

6) The number of iterations for each case is 10,000.

7) Compute the probability of the type I error by assigning the equivalent mean ratio. The probability of the type I error is defined as the ratio of number of times of rejection H_0 to number of iterations. When hypothesis testing at a significance level of 0.05, if the probability of type I error is between 0.025 – 0.075, the test statistics are considered to be capable of controlling the probability of type I error based on criteria from Bradley [1].

8) Compute the testing power by assigning the mean differences with the ratios of 1:1.5 for the small mean differences and 1:2 for the moderate mean differences. Power of the test is computed by the ratio of number of times of rejection H_0 to number of iterations.

4. Results

To evaluate the performance of seven test statistics, Z-test, t-test, Welch t-test, MWU, MAMWU, WMR, and BRW, we conducted the simulation data based on research method to calculate their probability of type I error and power of the test for testing the mean differences of two independent populations at the significance level of 0.05 ($\alpha = 0.05$) in which the populations have normal, exponential, log-normal, gamma, and Laplace distributions.

In comparison, the test statistics that have ability to control the probability of type I error and higher test estimation power are preferable to the choice of the test statistics that are robust.

Tables 1 to 5 display the common probability of type I error and testing power from simulation results classified by normal, exponential, log-normal, gamma, and Laplace distributions, respectively.

The following abbreviations are used in Tables. (n_1, n_2), $(V_1 : V_2)$, TE, PE 1:1.5 and PE 1:2 represent the sample sizes, the variance ratios, the probability of type I error, power of the test for the small and the moderate mean differences obtained from two populations, respectively.

Note bold entries indicate the test statistics that could control the probability of type I error, whereas bold and underlined entries indicate the test statistics that could control the probability of type I error while retaining the highest test estimation power.

According to normally distributed population in Table 1, it is clearly seen that Z-test, t-test, Welch t-test, MWU and WBR were all able to control the probability of type I error for all sample sizes in both equal and unequal variances.

Table 1 also indicates that for equal variance, Z-test had the highest testing power that could control the probability of type I error, as shown by the higher testing power relative to other test statistics for both small and moderate mean differences. For unequal variance with a small mean difference, Z-test had the highest test estimation power that had capacity for controlling the probability of type I error in most conditions, while t-test was quite similar in performance to MWU, which were close to 22.22% of the total conditions. When the mean difference was moderate, Z-test had the greatest estimation power.

Table 1. Probability of type I error and power of the test of seven test statistics for two independent populations with normal distributions

(n_1, n_2)	$(V_1:V_2)$	Z-test			t-test			Welch t-test		
		TE	PE1:1.5	PE1:2	TE	PE1:1.5	PE1:2	TE	PE1:1.5	PE1:2
(10, 10)	1:1	**0.0514**	**0.2023**	**0.6174**	**0.0516**	0.1855	0.5743	**0.0491**	0.1822	0.5689
	1:3	**0.0472**	**0.1285**	**0.3581**	**0.0530**	0.1218	0.3316	**0.0489**	0.1165	0.3204
	1:5	**0.0470**	0.1028	**0.2583**	**0.0553**	**0.1036**	0.2418	**0.0500**	0.0942	0.2220
	1:9	**0.0480**	0.0799	**0.1755**	**0.0576**	**0.0877**	0.1725	**0.0511**	0.0768	0.1535
(20, 20)	1:1	**0.0472**	**0.3635**	**0.8900**	**0.0478**	0.3503	0.8723	**0.0474**	0.3491	0.8715
	1:3	**0.0471**	**0.2094**	**0.6100**	**0.0493**	0.2021	0.5919	**0.0473**	0.1975	0.5852
	1:5	**0.0472**	**0.1544**	**0.4541**	**0.0504**	0.1509	0.4376	**0.0470**	0.1445	0.4263
	1:9	**0.0478**	0.1128	**0.3015**	**0.0508**	0.1139	0.2936	**0.0475**	0.1050	0.2807
(50, 50)	1:1	**0.0490**	**0.7255**	**0.9992**	**0.0501**	0.7174	0.9989	**0.0500**	0.7174	0.9989
	1:3	**0.0491**	**0.4396**	**0.9510**	**0.0501**	0.4314	0.9460	**0.0497**	0.4288	0.9454
	1:5	**0.0491**	**0.3182**	**0.8389**	**0.0527**	0.3130	0.8302	**0.0508**	0.3092	0.8273
	1:9	**0.0500**	0.2114	**0.6311**	**0.0527**	0.2098	0.6197	**0.0512**	0.2057	0.6138
(n_1, n_2)	$(V_1:V_2)$	MAMWU			MWU			BRW		
		TE	PE1:1.5	PE1:2	TE	PE1:1.5	PE1:2	TE	PE1:1.5	PE1:2
(10, 10)	1:1	0.9705	0.9833	0.9978	**0.0418**	0.1604	0.5207	0.8574	0.6429	0.9671
	1:3	0.9708	0.9757	0.9909	**0.0465**	0.1089	0.2919	0.8524	0.2572	0.8583
	1:5	0.9693	0.9757	0.9846	**0.0504**	0.0949	0.2191	0.8536	0.1856	0.6849
	1:9	0.9721	0.9724	0.9803	**0.0546**	0.0848	0.1616	0.9174	0.1697	0.5142
(20, 20)	1:1	0.9869	0.9966	1.0000	**0.0462**	0.3298	0.8519	0.7510	0.8168	0.8248
	1:3	0.9897	0.9943	0.9991	**0.0494**	0.1959	0.5587	0.7237	0.9222	0.7892
	1:5	0.9894	0.9930	0.9978	**0.0554**	0.1516	0.4152	0.7014	0.9467	0.8089
	1:9	0.9888	0.9909	0.9950	**0.0605**	**0.1246**	0.2889	0.6464	0.8098	0.8529
(50, 50)	1:1	0.9919	0.9996	1.0000	**0.0480**	0.6937	0.9989	0.4328	0.5317	0.3625
	1:3	0.9919	0.9990	1.0000	**0.0544**	0.4135	0.9290	0.2568	0.6760	0.5790
	1:5	0.9913	0.9974	0.9998	**0.0592**	0.3036	0.7971	0.2395	0.7758	0.7045
	1:9	0.9908	0.9957	0.9996	**0.0655**	**0.2180**	0.5893	0.1675	0.8590	0.7669
(n_1, n_2)	$(V_1:V_2)$	WBR								
		TE	PE1:1.5	PE1:2						
(10, 10)	1:1	**0.0478**	0.1727	0.5436						
	1:3	**0.0500**	0.1149	0.3026						
	1:5	**0.0527**	0.0986	0.2232						
	1:9	**0.0557**	0.0852	0.1632						
(20, 20)	1:1	**0.0462**	0.3298	0.8519						
	1:3	**0.0494**	0.1959	0.5587						
	1:5	**0.0551**	0.1516	0.4146						
	1:9	**0.0597**	0.1233	0.2871						
(50, 50)	1:1	**0.0485**	0.6964	0.9989						
	1:3	**0.0545**	0.4142	0.9293						
	1:5	**0.0593**	0.3037	0.7970						
	1:9	**0.0655**	**0.2180**	0.5891						

Table 2. Probability of type I error and power of the test of seven test statistics for two independent populations with exponential distribution

(n_1, n_2)	$(V_1:V_2)$	Z-test			t-test			Welch t-test		
		TE	PE1:1.5	PE1:2	TE	PE1:1.5	PE1:2	TE	PE1:1.5	PE1:2
(10, 10)	1:1	**0.0533**	**0.1417**	**0.2796**	**0.0449**	0.1162	0.2524	**0.0387**	0.1026	0.2233
	1:3	**0.0526**	**0.1223**	**0.3370**	**0.0675**	0.0875	0.3304	**0.0635**	0.0742	0.3052
	1:5	**0.0508**	**0.0993**	**0.2358**	**0.0841**	0.0639	0.1906	**0.0790**	0.0535	0.1631
	1:9	**0.0500**	**0.0803**	**0.1623**	**0.0983**	0.0640	0.1045	**0.0943**	0.0548	0.0828
(20, 20)	1:1	**0.0468**	**0.2312**	0.4971	**0.0429**	0.2180	**0.5177**	**0.0404**	0.2107	0.5058
	1:3	**0.0470**	**0.1914**	0.5856	**0.0581**	0.1604	**0.6169**	**0.0564**	0.1514	0.6085
	1:5	**0.0454**	**0.1447**	**0.4238**	**0.0689**	0.0996	0.4221	**0.0672**	0.0911	0.4074
	1:9	**0.0452**	**0.1049**	**0.2709**	**0.0786**	0.0663	0.2244	**0.0754**	0.0598	0.2068
(50, 50)	1:1	**0.0510**	0.4893	0.8920	**0.0495**	**0.4967**	**0.9239**	0.0494	0.4955	0.9229
	1:3	**0.0493**	0.3949	0.9465	**0.0547**	**0.3993**	**0.9592**	0.0539	0.3956	**0.9592**
	1:5	**0.0517**	**0.2814**	0.8149	**0.0586**	0.2607	**0.8588**	0.0573	0.2571	0.8571
	1:9	**0.0513**	**0.1915**	0.5816	**0.0624**	0.1483	**0.6090**	0.0610	0.1427	0.6017
(n_1, n_2)	$(V_1:V_2)$	MAMWU			MWU			BRW		
		TE	PE1:1.5	PE1:2	TE	PE1:1.5	PE1:2	TE	PE1:1.5	PE1:2
(10, 10)	1:1	0.9695	0.9794	0.9877	**0.0441**	0.1076	0.2321	0.6906	0.4289	0.4114
	1:3	0.9760	0.9760	0.9938	0.1022	0.0661	0.3783	0.3630	0.4502	0.4130
	1:5	0.9799	0.9725	0.9811	0.1337	0.0554	0.1443	0.3413	0.4507	0.3144
	1:9	0.9700	0.9719	0.9742	0.1586	0.0772	0.0698	0.3775	0.5861	0.2372
(20, 20)	1:1	0.9878	0.9947	0.9982	**0.0459**	0.1907	0.4505	0.8383	0.4080	0.4598
	1:3	0.9931	0.9927	0.9992	0.1636	0.0910	0.6825	0.5763	0.6947	0.3354
	1:5	0.9954	0.9907	0.9951	0.2209	0.0649	0.2561	0.5852	1.0000	0.5304
	1:9	0.9946	0.9909	0.9900	0.2661	0.1060	0.0868	0.5526	0.5257	0.6430
(50, 50)	1:1	0.9922	0.9988	0.9999	**0.0504**	0.4154	0.8345	0.5227	0.6369	0.5396
	1:3	0.9974	0.9938	1.0000	0.3299	0.1521	0.9754	0.7270	0.8551	0.4734
	1:5	0.9983	0.9927	0.9987	0.4464	0.0614	0.5392	0.7494	0.3002	0.6880
	1:9	0.9990	0.9941	0.9946	0.5274	0.1498	0.1142	0.7402	0.1080	0.6715
(n_1, n_2)	$(V_1:V_2)$	WBR								
		TE	PE1:1.5	PE1:2						
(10, 10)	1:1	**0.0501**	0.1183	0.2458						
	1:3	0.1054	0.0708	0.4017						
	1:5	0.1356	0.0578	0.1507						
	1:9	0.1596	0.0779	0.0716						
(20, 20)	1:1	**0.0459**	0.1907	0.4505						
	1:3	0.1635	0.0909	0.6825						
	1:5	0.2191	0.0649	0.2561						
	1:9	0.3641	0.1047	0.0860						
(50, 50)	1:1	**0.0511**	0.4175	0.8359						
	1:3	0.3299	0.1530	0.9756						
	1:5	0.4463	0.0614	0.5398						
	1:9	0.5268	0.1495	0.1142						

Table 3. Probability of type I error and power of the test of seven test statistics for two independent populations with log-normal distribution

(n_1, n_2)	$(V_1:V_2)$	Z-test			t-test			Welch t-test		
		TE	PE1:1.5	PE1:2	TE	PE1:1.5	PE1:2	TE	PE1:1.5	PE1:2
(10, 10)	1:1	**0.0530**	0.0572	0.0644	**0.0341**	0.0461	0.0723	**0.0276**	0.0370	0.0641
	1:3	**0.0503**	<u>0.0546</u>	<u>0.0611</u>	**0.0744**	0.0542	0.0438	**0.0695**	0.0495	0.0369
	1:5	**0.0483**	<u>0.0517</u>	<u>0.0574</u>	0.1008	0.0784	0.0635	0.0959	0.0740	0.0581
	1:9	**0.0451**	<u>0.0492</u>	<u>0.0538</u>	0.1273	0.1088	0.0902	0.1210	0.1020	0.0849
(20, 20)	1:1	**0.0531**	0.0576	0.0766	**0.0373**	0.0546	0.1053	**0.0338**	0.0515	0.1016
	1:3	**0.0500**	<u>0.0537</u>	<u>0.0634</u>	**0.0688**	0.0455	0.0429	**0.0669**	0.0432	0.0387
	1:5	**0.0487**	<u>0.0527</u>	<u>0.0599</u>	0.0868	0.0621	0.0470	0.0849	0.0600	0.0440
	1:9	**0.0455**	<u>0.0503</u>	<u>0.0599</u>	0.1047	0.0836	0.0666	0.1019	0.0818	0.0643
(50, 50)	1:1	**0.0528**	0.0720	0.1302	**0.0415**	0.0768	0.1717	**0.0406**	0.0760	0.1699
	1:3	**0.0492**	<u>0.0618</u>	<u>0.0932</u>	**0.0579**	0.0493	0.0708	**0.0566**	0.0482	0.0680
	1:5	**0.0480**	<u>0.0568</u>	<u>0.0790</u>	**0.0715**	0.0528	0.0524	**0.0706**	0.0521	0.0506
	1:9	**0.0471**	<u>0.0526</u>	<u>0.0661</u>	0.0799	0.0610	0.0519	0.0790	0.0596	0.0505

(n_1, n_2)	$(V_1:V_2)$	MAMWU			MWU			BRW		
		TE	PE1:1.5	PE1:2	TE	PE1:1.5	PE1:2	TE	PE1:1.5	PE1:2
(10, 10)	1:1	0.9688	0.9749	0.9834	**0.0466**	0.0681	0.1319	0.4867	0.8675	0.9479
	1:3	0.9831	0.9736	0.9716	0.1519	0.1044	0.0709	0.3452	0.4603	0.4643
	1:5	0.9824	0.9801	0.9760	0.2005	0.1600	0.1249	0.3800	0.6504	0.6440
	1:9	0.9845	0.9815	0.9780	0.2411	0.2084	0.1800	0.4258	0.7416	0.7091
(20, 20)	1:1	0.9895	0.9914	0.9944	**0.0499**	<u>0.1013</u>	<u>0.2622</u>	0.7965	0.1684	0.2100
	1:3	0.9960	0.9919	0.9934	0.2765	0.1707	0.0989	0.5806	0.6244	0.3818
	1:5	0.9965	0.9961	0.9920	0.3773	0.2860	0.2088	0.6018	0.8050	0.6160
	1:9	0.9977	0.9966	0.9959	0.4500	0.3793	0.3207	0.6534	0.8664	0.7791
(50, 50)	1:1	0.9899	0.9957	0.9996	**0.0511**	0.1970	0.5617	0.9749	0.3294	0.3279
	1:3	0.9996	0.9971	0.9945	0.5706	0.3410	0.1568	0.4491	0.6829	0.8675
	1:5	1.0000	0.9991	0.9979	0.7149	0.5699	0.4143	0.4832	0.7528	0.7615
	1:9	0.9998	0.9998	0.9995	0.8001	0.7128	0.6196	0.4901	0.7614	0.7676

(n_1, n_2)	$(V_1:V_2)$	WBR		
		TE	PE1:1.5	PE1:2
(10, 10)	1:1	**0.0508**	<u>0.0762</u>	<u>0.1407</u>
	1:3	0.1549	0.1079	0.0733
	1:5	0.2016	0.1619	0.1265
	1:9	0.2416	0.2089	0.1806
(20, 20)	1:1	**0.0499**	<u>0.1013</u>	<u>0.2622</u>
	1:3	0.2756	0.1705	0.0989
	1:5	0.3751	0.2840	0.2073
	1:9	0.4452	0.3765	0.3172
(50, 50)	1:1	**0.0522**	<u>0.1985</u>	<u>0.5630</u>
	1:3	0.5706	0.3411	0.1570
	1:5	0.7146	0.5699	0.4131
	1:9	0.7991	0.7122	0.6192

Table 4. Probability of type I error and power of the test of seven test statistics for two independent populations with gamma distribution

(n_1, n_2)	$(V_1 : V_2)$	Z-test			t-test			Welch t-test		
		TE	PE1:1.5	PE1:2	TE	PE1:1.5	PE1:2	TE	PE1:1.5	PE1:2
(10, 10)	1:1	**0.0460**	0.0483	0.2875	**0.0405**	0.2252	0.6156	**0.0347**	0.2180	0.6071
	1:3	**0.0468**	0.7080	**0.9819**	**0.0640**	**0.7411**	0.9730	**0.0617**	0.7272	0.9701
	1:5	**0.0511**	**0.9373**	**0.9996**	0.0804	0.9075	0.9977	0.0765	0.8939	0.9970
	1:9	**0.0463**	**0.9972**	**1.0000**	0.0863	0.9841	1.0000	0.0822	0.9772	1.0000
(20, 20)	1:1	**0.0530**	0.1121	0.6524	**0.0470**	0.3728	0.8637	**0.0449**	0.3708	0.8617
	1:3	**0.0475**	0.9568	**1.0000**	**0.0585**	**0.9601**	0.9998	**0.0573**	0.9594	0.9997
	1:5	**0.0492**	**0.9984**	**1.0000**	0.0650	0.9973	**1.0000**	0.0633	0.9973	**1.0000**
	1:9	**0.0501**	**1.0000**	**1.0000**	0.0768	1.0000	1.0000	**0.0728**	**1.0000**	**1.0000**
(50, 50)	1:1	**0.0492**	0.3879	0.9870	**0.0446**	0.6976	0.9966	**0.0441**	0.6968	0.9965
	1:3	**0.0509**	**0.9999**	**1.0000**	**0.0535**	**0.9999**	**1.0000**	**0.0531**	**0.9999**	**1.0000**
	1:5	**0.0486**	**1.0000**	**1.0000**	**0.0544**	**1.0000**	**1.0000**	**0.0536**	**1.0000**	**1.0000**
	1:9	**0.0475**	**1.0000**	**1.0000**	0.0626	**1.0000**	**1.0000**	0.0613	**1.0000**	**1.0000**

(n_1, n_2)	$(V_1 : V_2)$	MAMWU			MWU			BRW		
		TE	PE1:1.5	PE1:2	TE	PE1:1.5	PE1:2	TE	PE1:1.5	PE1:2
(10, 10)	1:1	0.9715	0.9917	0.9991	**0.0402**	0.3053	0.7143	**0.0277**	**0.4030**	0.5026
	1:3	0.9770	0.9998	1.0000	0.0850	0.8150	0.9842	0.7546	0.5957	0.5021
	1:5	0.9770	1.0000	1.0000	0.1143	0.9462	0.9987	0.8519	0.7098	0.5539
	1:9	0.9761	1.0000	1.0000	0.1313	0.9939	1.0000	0.8988	0.7645	0.8976
(20, 20)	1:1	0.9895	0.9994	1.0000	**0.0445**	**0.5811**	**0.9633**	0.1572	0.6635	0.0813
	1:3	0.9941	1.0000	1.0000	0.1364	0.9904	1.0000	0.2758	0.6879	0.7086
	1:5	0.9935	1.0000	1.0000	0.1930	0.9996	1.0000	0.8581	0.7876	0.7042
	1:9	0.9939	1.0000	1.0000	0.2388	1.0000	1.0000	0.2721	0.5236	0.8516
(50, 50)	1:1	0.9911	1.0000	1.0000	**0.0469**	0.9305	**1.0000**	0.4715	0.1544	0.4084
	1:3	0.9972	1.0000	1.0000	0.2715	1.0000	1.0000	0.9575	0.1577	0.1761
	1:5	0.9985	1.0000	1.0000	0.3915	1.0000	1.0000	0.9579	0.8291	0.6487
	1:9	0.9987	1.0000	1.0000	0.4841	1.0000	1.0000	0.6214	0.6881	0.7045

(n_1, n_2)	$(V_1 : V_2)$	WBR		
		TE	PE1:1.5	PE1:2
(10, 10)	1:1	**0.0454**	0.3171	**0.7194**
	1:3	0.0886	0.8242	0.9850
	1:5	0.1171	0.9496	0.9987
	1:9	0.1325	0.9944	1.0000
(20, 20)	1:1	**0.0445**	**0.5811**	**0.9633**
	1:3	0.1364	0.9904	1.0000
	1:5	0.1929	0.9996	1.0000
	1:9	0.2365	1.0000	1.0000
(50, 50)	1:1	**0.0477**	**0.9306**	**1.0000**
	1:3	0.2722	1.0000	1.0000
	1:5	0.3915	1.0000	1.0000
	1:9	0.4840	1.0000	1.0000

Table 5. Probability of type I error and power of the test of seven test statistics for two independent populations with Laplace distribution

(n_1, n_2)	$(V_1:V_2)$	Z-test			t-test			Welch t-test		
		TE	PE1:1.5	PE1:2	TE	PE1:1.5	PE1:2	TE	PE1:1.5	PE1:2
(10, 10)	1:1	0.1621	0.1493	0.3575	**0.0441**	0.1187	0.3114	**0.0427**	0.1098	0.2951
	1:3	**0.0375**	0.0696	0.1797	**0.0435**	0.0930	0.2319	**0.0410**	0.0893	0.2234
	1:5	0.0231	0.0421	0.1112	**0.0441**	0.0815	0.1878	**0.0416**	0.0784	0.1809
	1:9	0.0175	0.0281	0.0590	**0.0457**	0.0734	0.1454	**0.0423**	0.0683	0.1386
(20, 20)	1:1	0.1667	0.2277	0.5939	**0.0490**	0.1814	0.5249	**0.0484**	0.1747	0.5144
	1:3	**0.0411**	0.1015	0.3454	**0.0516**	0.1357	0.3974	**0.0511**	0.1338	0.3943
	1:5	0.0233	0.0648	0.2225	**0.0472**	0.1194	0.3326	**0.0460**	0.1183	0.3302
	1:9	0.0181	0.0365	0.1223	**0.0530**	0.0984	0.2524	**0.0513**	0.0962	0.2484
(50, 50)	1:1	0.1657	0.4244	0.9202	**0.0478**	0.3571	0.8767	**0.0477**	0.3550	0.8748
	1:3	**0.0354**	0.2155	0.7202	**0.0492**	0.2628	0.7543	**0.0489**	0.2622	0.7535
	1:5	0.0243	0.1407	0.5422	**0.0476**	0.2210	0.6509	**0.0476**	0.2207	0.6506
	1:9	0.0151	0.0704	0.3220	**0.0504**	0.1618	0.5153	**0.0498**	0.1610	0.5132

(n_1, n_2)	$(V_1:V_2)$	MAMWU			MWU			BRW		
		TE	PE1:1.5	PE1:2	TE	PE1:1.5	PE1:2	TE	PE1:1.5	PE1:2
(10, 10)	1:1	0.9708	0.9787	0.9924	**0.0420**	0.1315	0.3466	0.2243	0.8595	0.7424
	1:3	0.9713	0.9771	0.9893	**0.0414**	<u>0.0986</u>	0.2479	0.5548	0.4996	0.3286
	1:5	0.9696	0.9742	0.9859	**0.0406**	0.0827	0.2014	**<u>0.0600</u>**	**<u>0.5193</u>**	**<u>0.4979</u>**
	1:9	0.9705	0.9750	0.9844	**0.0424**	0.0705	0.1615	0.0751	0.0040	0.8783
(20, 20)	1:1	0.9899	0.9958	0.9993	**0.0498**	**<u>0.2407</u>**	**<u>0.6506</u>**	0.1743	0.8195	0.7446
	1:3	0.9900	0.9930	0.9978	**0.0511**	0.1652	**<u>0.4874</u>**	0.8801	0.2660	0.1811
	1:5	0.9884	0.9947	0.9977	**0.0506**	**<u>0.1396</u>**	0.4130	0.8206	0.5149	0.9316
	1:9	0.9891	0.9933	0.9957	**0.0541**	0.1141	0.3181	0.3221	0.8843	0.8395
(50, 50)	1:1	0.9919	0.9986	0.9999	**0.0505**	**<u>0.5011</u>**	**<u>0.9620</u>**	0.3898	0.9042	0.1275
	1:3	0.9928	0.9981	1.0000	**0.0506**	**<u>0.3571</u>**	0.8760	0.3926	0.0746	0.4801
	1:5	0.9907	0.9969	0.9999	**0.0513**	0.2882	**<u>0.7937</u>**	0.2306	0.3321	0.6598
	1:9	0.9922	0.9954	0.9997	**0.0524**	**<u>0.2263</u>**	0.6549	0.2846	0.3057	0.9978

(n_1, n_2)	$(V_1:V_2)$	WBR		
		TE	PE1:1.5	PE1:2
(10, 10)	1:1	**<u>0.0440</u>**	**<u>0.1420</u>**	**<u>0.3620</u>**
	1:3	**0.0420**	0.0970	**<u>0.2790</u>**
	1:5	**0.0590**	0.0950	0.2330
	1:9	**0.0400**	**<u>0.0790</u>**	**<u>0.1700</u>**
(20, 20)	1:1	**0.0440**	0.2320	0.6430
	1:3	**0.0500**	**<u>0.1710</u>**	0.4790
	1:5	**0.0450**	0.1360	**<u>0.4240</u>**
	1:9	**0.0670**	**<u>0.1190</u>**	**<u>0.3230</u>**
(50, 50)	1:1	**0.0470**	0.4940	0.9600
	1:3	**0.0490**	0.3520	**<u>0.8830</u>**
	1:5	**0.0380**	**<u>0.3070</u>**	0.7810
	1:9	**0.0380**	0.2180	**<u>0.6600</u>**

As shown in Table 2, when the populations had exponential distribution, it was found that Z-test, t-test, and Welch t-test were capable of controlling the probability of type I error for all sample sizes in both equal and unequal variances. On the other hand, MWU and WBR were capable of controlling the probability of type I error for only equal variance.

Additionally, for equal variance, the test statistics with the highest testing power for controlling the probability of type I error in most conditions were Z-test for small mean difference and t-test for moderate mean difference. For unequal variance with small mean difference, the test statistics with the highest testing power for controlling the probability of type I error were Z-test and t-test. Furthermore, when the mean difference was moderate, the test statistics with the highest testing power for controlling the probability of type I error were Z-test, followed by t-test, Welch t-test, respectively.

Based on log-normal distribution given in Table 3, the results demonstrated that Z-test had the ability to control the probability of type I error for all sample sizes in both equal and unequal variances. For equal variance, t-test, Welch t-test, MWU and WBR had the ability to control the probability of type I error for all sample sizes. Nevertheless, for unequal variance, t-test and Welch t-test with unequal variance ratio of 1:3, and Welch t-test for $(n_1, n_2) = (50, 50)$ with unequal variance ratio of 1:5 were able to control the probability of type I error.

This Table 3 confirms that for equal variance with small and moderate mean differences, WBR had superior test estimation power for controlling the probability of type I error, followed by MWU. For unequal variance with small and moderate mean differences, Z-test was very effective with the highest testing power that could control the probability of type I error.

Table 4 gives the results of gamma distribution. It was found that Z-test could control the probability of type I error for all sample sizes in both equal and unequal variances. Moreover, t-test and Welch t-test could also control the probability of type I error for all conditions except for $(n_1, n_2) = (10, 10)$ with unequal variance ratios of 1:5 and 1:9. For equal variance, MWU and WBR could control the probability of type I error for all sizes, while BRW could control the probability of type I error for only $(n_1, n_2) = (10, 10)$.

Based on power of the test for equal variance with small and moderate mean differences, WBR dominated the other test statistics with the highest test estimation power and ability to control the probability of type I error. Whereas the second largest test estimation power and ability to control the probability of type I error were MWU and BRW in the case of small mean difference, and MWU in the case of moderate mean difference. For unequal variance, Z-test was suitable for both small and moderate mean differences. While the second most powerful test statistics were t-test for small mean difference, and Welch t-test for moderate mean difference.

Table 5 displays the results of Laplace distribution, where it was observed that t-test, Welch t-test, MWU and WBR were able to control the probability of type I error for all cases in

both equal and unequal variances. For unequal variance, Z-test was able to control the probability of type I error with variance ratio of 1:3, while BRW could control the probability of type I error for only $(n_1, n_2) = (10, 10)$ with variance ratio of 1:5.

Considering power of the test for equal variance with small and moderate mean differences, the superior test statistics with the highest testing power for controlling the probability of type I error in most conditions were MWU and WBR, respectively. For unequal variance with small mean differences, MWU and WBR had the greatest estimation power for controlling the probability of type I error, followed by BRW. For moderate mean difference, the highest testing power which could control the probability of type I error were WBR, MWU and BRW, respectively.

5. Conclusion and Discussion

The results showed that when the distributions of two populations assumed to be normal, exponential, log-normal, gamma and Laplace distributions in the case of homogeneous variances, Z-test, t-test, Welch t-test, MWU and WBR yielded acceptable capability of controlling the probability of type I error for all sample sizes based on Bradley's criteria except for Z-test, which was not suitable for Laplace distribution. Consequently, the test statistics were robust to changes in the distribution of the population. Regardless of heterogeneity of variances for two population groups, all test statistics yielded a lower acceptable capability of controlling the probability of type I error. Z-test, t-test, and Welch t-test had the capability of controlling the probability of type I error for almost conditions under all considered distributions. Moreover, MWU and WBR were superior for all sample sizes under normal and Laplace distributions, whereas BRW was appropriate for gamma distribution with only $(n_1, n_2) = (10, 10)$ and unequal variance ratio of 1:9. MAMWU performing poorly for all conditions.

For equal variance with both small and moderate mean differences, the test statistics with the highest estimation power for controlling the probability of type I error in most conditions were Z-test for normal distribution, WBR for log-normal and gamma distributions, and MWU for Laplace distribution. Z-test was better for small mean difference, while t-test was superior for moderate mean difference when two populations had exponential distributions.

For unequal variance with both small and moderate mean differences, Z-test had the highest testing power and ability to control the probability of type I error in most conditions for normal, exponential, log-normal and gamma distributions. MWU and WBR were superior for Laplace distribution when the mean difference was small. WBR was suitable when the mean difference was moderate.

Besides, it was found that MAMWU gave the highest test estimation power for every condition but did not have the capacity to control the probability of type I error. This related to

the research of Eriobu and Umeh [4]. When two populations had gamma distributions and equal variance, WBR was recommended for testing hypothesis because this test statistic had the highest testing power for controlling the probability of type I error. This is consistent with the work of Sangthong and Klubnual [11]. Furthermore, the results also revealed that the testing power estimation tended to be higher as sample sizes increased when the variance ratios were identical between two populations.

Conflicts of Interest: The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

[1] J.V. Bradley, Robustness?, Br. J. Math. Stat. 31 (1978), 144-152. https://doi.org/10.1111/j.2044-8317.1978.tb00581.x.

[2] D. Wongchai, M. Lao, A. Thongteeraparp, A Comparison of Parametric and Nonparametric Test Statistics for Testing the Mean Difference of Two Independent Populations, Thai Sci. Technol. J. 31 (2023), 1-13. https://doi.org/10.14456/TSTJ.2023.11.

[3] B. Efron, R.J. Tibshirani, An Introduction to the Bootstrap, Chapman & Hall, New York, (1993).

[4] N.O. Eriobu, E.U. Umeh, Evaluation of Sensitivity of Three Selected Two-sample Nonparametric Tests, J. Basic Phys. Res. 9 (2019), 45-56.

[5] M.W. Fagerland, L. Sandvik, Performance of Five Two-Sample Location Tests for Skewed Distributions with Unequal Variances, Contemp. Clinic. Trials. 30 (2009), 490-496. https://doi.org/10.1016/j.cct.2009.06.007.

[6] P. Mann, Introductory Statistics, 7th ed., John Wiley & Sons, New Jersey, (2010).

[7] D. C. Montgomery, G.C. Runger, Applied Statistics and Probability for Engineers, 5th ed., John Wiley & Sons, New York, (2011).

[8] S. Montri, Nonparametric Bootstrap Method for Location Testing between Two Populations under Combined Assumption Violations, Burapha Sci. J. 25 (2020), 864-879.

[9] I. C. Oyeka, U. M. Okeh, Modified Intrinsically Ties Adjusted Mann-Whitney U Test, IOSR J. Math. 7 (2013), 52-56.

[10] J. Reiczigel, I. Zakariás, L. Rózsa, A Bootstrap Test of Stochastic Equality of Two Populations, Amer. Statistician. 59 (2005), 156-161. https://doi.org/10.1198/000313005x23526.

[11] M. Sangthong, P. Klubnual, Performance of Seven Statistics for Mean Difference Testing Between Two Populations Under Combined Assumption Violations, Naresuan Univ. J.: Sci. Technol. 29 (2021), 112-126. https://doi.org/10.14456/nujst.2021.40.

[12] S. Thidarat, Y. Preeya, S. Ratiwat, R. Wirayut, S. Saichon, Comparison of Nonparametric Test Statistics for Testing Population Mean between Two Independent Populations, Thai Sci. Technol. J. 26 (2018), 739-751.

[13] W. Wanishsakpong, K. Sodrung, A. Thongteeraparp, A Comparison of Nonparametric Statistics and Bootstrap Methods for Testing Two Independent Populations with Unequal Variance, Int. J. Anal. Appl. 21 (2023), 36. https://doi.org/10.28924/2291-8639-21-2023-36.

[14] B.L. Welch, The Significance of the Difference between Two Means When the Population Variables are Unequal, Biometrika. 20 (1938), 350-362. https://doi.org/10.2307/2332010.

[15] F. Wilcoxon, Individual Comparisons by Ranking Methods, Biometrics. 1 (1945), 80-83.

[16] J.C.F. Winter, Using the Student's T-Test With Extremely Small Sample Sizes, Pract. Assess. Res. Eval. 18 (2013), 1-12. https://doi.org/10.7275/e4r6-dj05.