Sains Malaysiana 51(11)(2022): 3829-3841
http://doi.org/10.17576/jsm-2022-5111-26

A Comparison of Efficiency of Test Statistics for Detecting Outliers in Normal Population

(Suatu Perbandingan Kecekapan Ujian Statistik untuk Mengesan Maklumat Tepian dalam Populasi Normal)

Kullaphat Promtep[1,2] Phontita Thiuthad[1,2,*] & Natchita Intaramo[2]

[1]Statistics and Applications Research Unit, Faculty of Science, Prince of Songkla University, Hat Yai, Songkhla, 90110 Thailand
[2]Division of Computational Science, Faculty of Science, Prince of Songkla University, Hat Yai, Songkhla, 90110 Thailand

Received: 20 October 2021/Accepted: 24 June 2022

ABSTRACT

The objective of this research was to compare the efficiency among the test statistics which are used to detect outliers by testing hypothesis methods. The test statistics considered were Dixon's test, Ferguson's test, Grubbs' test, T_w-test, and Tietjen-Moore's test. The outliers were divided, by how far they are, into two groups: mild and extreme outliers. The efficiency of the test statistics was measured by the probability of type I error and the power of the test. The results showed that Tietjen-Moore's test can control the probability of type I error according to Cochran and Bradley criteria in every situation. T_w-test has highest sensitivity in detecting one outlier when the sample size is small or moderate but, if the sample size is large, Grubbs' test performs better. In the case of detecting one extreme outlier, the power of four tests tend to increase as the sample size increases at the significance level 0.01. Given that k outliers are detected, Tietjen-Moore's test provides higher power than T_w-test when k equals 10% of sample size when the outliers are both mild and extreme, contrary to the case when k make up for 20%.

Keywords: Detection of outliers; normal distribution; power of the test; Tietjen-Moore's test; type I error

ABSTRAK

Objektif kajian ini adalah untuk membandingkan kecekapan antara statistik ujian yang digunakan untuk mengesan maklumat tepian dengan menguji kaedah hipotesis. Statistik ujian yang dipertimbangkan ialah ujian Dixon, ujian Ferguson, ujian Grubbs, ujian T_w dan ujian Tietjen-Moore. Maklumat tepian dibahagikan mengikut jarak kepada dua kumpulan: maklumat tepian ringan dan maklumat tepian melampau. Kecekapan ujian statistik diukur dengan kebarangkalian ralat jenis I dan kuasa ujian. Keputusan menunjukkan bahawa ujian Tietjen-Moore boleh mengawal kebarangkalian ralat jenis I mengikut kriteria Cochran dan Bradley dalam setiap situasi. Ujian T_w mempunyai kepekaan tertinggi dalam mengesan satu maklumat tepian apabila saiz sampel kecil atau sederhana tetapi jika saiz sampel besar, ujian Grubbs menunjukkan prestasi yang lebih baik. Dalam kes mengesan satu maklumat tepian melampau, kuasa empat ujian cenderung meningkat apabila saiz sampel meningkat pada tahap keertian 0.01. Memandangkan k maklumat tepian dikesan, ujian Tietjen-Moore memberikan kuasa yang lebih tinggi daripada ujian T_w apabila k bersamaan dengan 10% saiz sampel apabila maklumat tepian adalah ringan dan melampau, bertentangan dengan kes apabila k membentuk 20%.

Kata kunci: Kuasa ujian; pengesan maklumat tepian; ralat jenis I; taburan normal; ujian Tietjen-Moore

Introduction

Applying statistical knowledge and analysis for solving a problem is very important to improve the quality of data and make it more reliable but dealing with the data is a challenge for researchers. One of the data management/cleaning problems is that there are outliers in the collected data. The outliers are extreme values that deviate from the other values in the abnormal distance (Hawkins 1980). In general, the outliers are between the inner and outer fences, shown in Figure 1, and can be calculated as follows.

- Lower inner fence: $Q_1 - 1.5IQR$
- Upper inner fence: $Q_3 + 1.5IQR$
- Lower outer fence: $Q_1 - 3IQR$
- Upper outer fence: $Q_3 + 3IQR$

A point beyond an inner fence on either side is considered as a mild outlier and a point beyond an outer fence is considered as an extreme outlier.

In the data collection process, we may capture some inaccurate information which are the outliers. If there are some outliers in the data set, they may affect the mean of the data or may affect the distribution of the data. This can lead the statistician to pick up an incorrect statistical tool for analysis, resulting in a lack of quality and reliability, which in turn affect the research work could not be used properly.

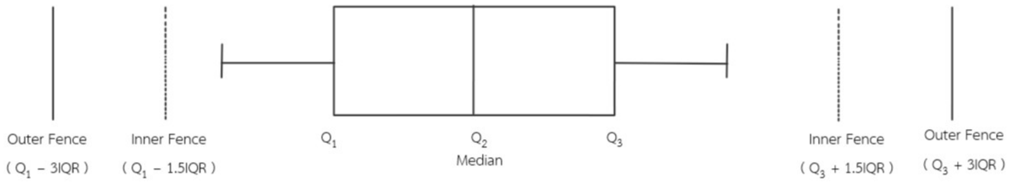


FIGURE 1. Box plot

Outlier detection is an interesting problem to study. There are various statistical methods for detecting the outliers in the univariate data. For example, one can detect the outliers by the box plot method, which is an easy one, but using this method may give an unclear result and it depends on many reasons such as the experience of the analyst. The other way to detect the outlier is by using a statistical hypothesis test. It is a way to determine whether the data has an outlier, which has been studied and developed by many statisticians.

RELATED WORKS

Rattanaloetnusorn (1991) studied three methods, which are Tietjen Moore and Beckman method (TMB), Mervyn G. Marasinghe (M) method, and G. Barrie Wetherill (GB) method, for detecting outliers in the simple linear regression analysis with a view study of two residual distributions which are a heavy-tailed distribution (such as scale contaminated normal, location contaminated normal and t-distributions) and a right-handed skew distribution (such as lognormal, gamma and Weibull distributions) when there are one, two, and three outliers in the sample. The probability of type I error and the power of the test statistic were compared in this research. The results showed that the TMB method can control the probability of type I error less than the other two methods. GB test gives the highest power when there is one outlier in the sample but, when there are two or three outliers, the M test has the highest power. In the overall conclusion, the two distributions give us the same results.

Efstathiou (2006) studied the efficiency of Dixon's test (Dixon 1953) since the critical values used in the package may be outdated. He compared the probability of type I error of Dixon's test when the sample sizes are 3, 4, ..., 30 at significant levels 0.20, 0.10, 0.05, 0.04, 0.02, and 0.01. The result showed that the test is accurate when the sample size is very small that is only a sample size of three or four.

Patchayaluck (2013) estimated the probability of type I error and the power of three test statistics for detecting an outlier by using stochastic and classical approaches. The test statistics considered were Dixon's test (Dixon 1953), Grubbs' test (Grubbs 1969), and Tietjen-Moore's test (Tietjen & Moore 1972). The critical values of Dixon's test and Grubbs' test using stochastic and classical approaches are closed in all cases. Tietjen-Moore's test using the stochastic approach has higher critical values than those of classical one in all situations. All of the tests can control the probability of type I error in all cases except Tietjen-Moore's test using stochastic approach cannot control the probability of type I error in all cases.

Jareankam (2013) proposed the test statistics (T_{P1} and T_{P2}) for detecting outliers based on Ferguson's test (T_{N14} and T_{N15}) which was calculated by skewness coefficient and kurtosis coefficient (Ferguson 1961). The efficiency of the T_{P1} and T_{P2} test statistics were compared with the T_{N14} and T_{N15}. The study is divided into two cases: detecting the outliers on right by T_{P1} and T_{N14}, and detecting the outliers on two sides of the sample by T_{P2} and T_{N15}. The results of this study showed that T_{P1} can control the probability of type I error in all situations while T_{N14} can control only when the sample sizes are small. T_{N15} has the power of the test better than T_{P2} when the sample size is quite small but, with the large or very large sample sizes, the efficiency of T_{P2} and T_{N15} seems the same.

Rahman, Sathik and Kannan (2014) used three different methods to detect outliers in three variables which are Grubbs' method, inner and outer fence rule method, and three-sigma rule method to find out how many outliers were detected for each variable from each method. The results showed that the inner fence rule method has the highest sensitivity in detecting outliers. Grubbs' method gave the results close to the results from the outer fence rule method.

Jareankam (2020) studied the detection of outliers and proposed T_W-test which was developed from the concept idea of the generalized extreme Studentized deviate (GESD) test by Rosner (1975). The probability of type I error and the power of the test were considered under a simulation of normal distributions with 1000 replications at significance level 0.05. In this research, the probability of type I error was controlled by Cochran's criterion (Cochran 1954) in every situation. The results when there are *k* outliers in the sample showed that the percentage of correct decision are greater than 95 in every situation.

As we have reviewed several different ways of outlier detection, therefore, the test statistics for detecting outliers, considered in this research, are Dixon's test, Ferguson's test, Grubbs' test, Tietjen-Moore's test, and T_W-test. The objective of this research was to consider the probability of type I error controlled by the criteria of Cochran and Bradley and to compare the power of the test statistics for detecting one outlier and k ($\geq$ 2) outliers in the normal population. The materials, methods, simulation process, results and discussion, and conclusion of this research will be provided as follows.

Materials and Methods

We compare the efficiency of all test statistics for detecting outliers in a normal population through a simulation study (the process of simulation will be provided in the next section). The efficiency is in terms of the probability of type I error and the power of the tests. The test statistics can control the probability of type I error based on Cochran's criteria (Cochran 1954) if they are within the following ranges:

$\hat{\alpha}$ is in the range (0.007, 0.015) at the significance level 0.01,

$\hat{\alpha}$ is in the range (0.04, 0.06) at the significance level 0.05,

$\hat{\alpha}$ is in the range (0.081, 0.119) at the significance level 0.1.

The test statistics can control the probability of type I error based on Bradley's criterion (Bradley 1978) if they are within the following ranges:

$\hat{\alpha}$ is in the range (0.005, 0.015) at the significance level 0.01,

$\hat{\alpha}$ is in the range (0.025, 0.075) at the significance level 0.05,

$\hat{\alpha}$ is in the range (0.05, 0.15) at the significance level 0.1.

The test statistics are divided into two cases, i.e., the ones for detecting one outlier and the other ones for detecting k ($\geq$ 2) outliers which k are considered into two cases, that is, 10% and 20% of sample size *n*. The outliers, in this research, are also divided into two groups, i.e.,

- mild outlier which is in the interval ($Q_3 + 1.5IQR$, $Q_3 + 3IQR$),
- extreme outlier which is in the interval ($Q_3 + 3IQR$, $Q_3 + 4.5IQR$),

which we have considered only when the outliers are on the right side of the distribution. In the following, we will provide the materials for this research. It is divided into two subsections.

Detecting One Outlier

This subsection provides the test statistics for detecting one outlier which are defined for the following hypotheses:

H_0: There is no outlier in the population.

H_1: There is an outlier on the right side of the mean.

Reject H_0 at significance level α if each test statistic value is greater than its critical value.

Dixon's Test

Dixon (1953) offered a test for detecting an outlier in small sample sizes. The statistic is calculated by a different of a suspected outlier and the value that is closest to the suspected one divided by the range of the data as the following formula:

$$\text{Dixon's test} = \frac{x_{(n)} - x_{(n-1)}}{x_{(n)} - x_{(1)}}$$

where the order statistics $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$ are defined by sorting the data $x_1, \ldots, x_n$ of the sample size n in ascending order. Critical values can be found in the table of Dixon's critical values (Dixon 1951; Verma & Quiroz-Ruiz 2006).

Ferguson's Test

Ferguson (1961) proposed the coefficient of skewness and the coefficient of kurtosis to be a statistic for detecting an outlier with the assumption that the data follows a normal distribution. The test statistic is used to determine whether a minimum or maximum observation value is an outlier which can be computed as follows.

$$\text{Ferguson's test} = \frac{\sqrt{n}\sum_{i=1}^{n}(x_i - \bar{x})^3}{\left[\sum_{i=1}^{n}(x_i - \bar{x})^2\right]^{3/2}} \equiv \frac{M_3}{M_2^{3/2}}$$

where $M_r = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^r$ is the rth sample moment about the mean $\bar{x}$ and $r = 2, 3$. Critical values can be found in Ferguson's critical value table (Barnett & Lewis 1984).

Grubbs' Test

Grubbs (1969) proposed a hypothesis testing for detecting a single outlier in a univariate data set that follows an approximately normal distribution. The statistic is calculated by a difference between the suspected outlier and the mean of the sample divided by the standard deviation which is calculated from all data including the outlier as follows.

$$\text{Grubbs' test} = \frac{|\bar{x} - x_i|}{s}$$

where $\bar{x} = \frac{1}{n}\sum_{i=1}^{n}x_i, s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$ and x_i is a suspected outlier. Critical values can be found in the Grubbs' critical value table (Grubbs 1969).

Detecting k (≥ 2) Outliers

The subsection provides the test statistics for detecting more than one outlier ($k \geq 2$) which are defined for the hypotheses:

H_0: There is no outlier in the population.

H_1: There are k outliers on the right side of the mean.

Tietjen-Moore's Test

Tietjen and Moore (1972) has developed Grubbs' test to be able to detect multiple outliers with the assumption that the data follows a normal distribution. The developed test will become Grubbs' test if there is only one outlier in the sample. The formula of the test is provided as follows.

$$\text{Tietjen-Moore's test} = \frac{\sum_{i=1}^{n-k}(x_i - \bar{x}_k)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

where $\bar{x}_k = \frac{1}{n-k}\sum_{i=1}^{n-k}x_i$ and $\bar{x}$ is the sample mean of size n. Critical values are determined by Monte Carlo simulation. The simulation is typically performed by generating 10,000 standard normal random samples of size n and computing the Tietjen-Moore test. The statistical value obtained from the data is compared to the reference approximate distribution which is between zero and one. If there are some outliers in the sample, the test statistic is close to zero. Otherwise, there are no outliers in the sample.

T_W-Test

T_W-test is developed from the concept idea of GESD test by Rosner (1975) which fixed the problem of the outlier detection using Grubbs' test that one or two outliers at a time may cause more error. T_W-test, therefore, may be used to test whether there is one or more than one outlier in the sample. This test statistic is under an assumption of a normally distributed population and calculated by Jareankam (2020)

$$T_W\text{-test} = \frac{\frac{1}{ns^3}\sum_{i=1}^{n}(x_i - \bar{x})^3}{\sqrt{\left(\frac{n-1}{n}\right)^3 \frac{6(n-2)}{(n+1)(n+3)}}} \xrightarrow{d} N(0,1)$$

where $\bar{x} = \frac{1}{n}\sum_{i=1}^{n}x_i$ and $s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$.

SIMULATION STUDY

The simulation processes of this research are provided in this section. The data are simulated from a normal distribution with zero mean and one variance, the sample sizes of 5, 10, 15, 20, 30, 50, and 100 with 10^4 replications for each situation by the R program version 4.1.2. The significant levels used are 0.01, 0.05, and 0.1.

PROBABILITY OF TYPE I ERROR

The computations of the probability of type I error are done by the following iterative processes: 1. Generate a normal random sample of size n with zero mean and one variance. 2. Compute all test statistics. 3. Compare the test statistical values from step two with its critical regions to decide whether reject or accept the null hypothesis (H_0) at a significance level α. 4. Repeat steps one to three for 10^4 replications and record how many times the null hypothesis is rejected at a significance level α. 5. Calculate the estimation of the probability of type I error for each test statistic:

$$\hat{\alpha} = \frac{\text{number of } \{\text{H}_0 \text{ is rejected}\} \text{ when H}_0 \text{ is true}}{10000}$$

Power of the Test

The computations of the power of the test, when there are one and $k = 10\%n$, $20\%n$ outliers in the sample, are done by the following iterative processes:1. Generate a normal random sample of size $(n - 1)$ or $(n - k)$ with zero mean and one variance. 2. Select (an) outlier(s) by a simple random sampling method from the following intervals divided by two types of outliers: a) mild outlier: $(Q3 + 1.5IQR, Q3 + 3IQR)$ and b) extreme outlier: $(Q3 + 3IQR, Q3 + 4.5IQR)$. 3. Calculate the test statistics for detecting one outlier which are Dixon's test, Ferguson's test, Grubbs' test, and Tietjen-Moore's test and calculate the test statistics for detecting k outliers which are Tietjen-Moore's test and T_W-test. 4. Compare the test statistical values from step three with its critical regions to decide whether reject or accept the null hypothesis (H_0) at a significance level α. 5. Repeat steps one to four for 10^4 replications and record how many times the null hypothesis is rejected at a significance level α. 6. Calculate the estimation of the power of each test statistic:

$$\widehat{1-\beta} = \frac{\text{number of } \{\text{H}_0 \text{ is rejected}\} \text{ when H}_0 \text{ is false}}{10000}$$

R code for the statistical analyses done are available at https://github.com/KullaphatP/R-code.

RESULTS AND DISCUSSION

For our simulation study to compare the performance of Dixon's test, Ferguson's test, Grubbs' test, Tietjen-Moore's test, and T_W-test in terms of type I error, results of which are presented in Table 1 and Figure 2. We can see that Tietjen-Moore's test can control the probability of type I error by Cochran and Bradley Criteria in every situation as considered but Dixon's test can only control it by both criteria when $n \leq 30$. The T_W-test cannot control type I error when the sample sizes are small such as $n = 5$ and 10 at significance level 0.01. Ferguson's test can control the probability of type I error by Bradley Criteria when $n \leq 30$ at significance levels 0.01 and 0.05 and by Cochran Criteria when $n \leq 30$ at significance levels 0.01 but not when $n = 15$ at significance level 0.05. While Grubbs' test can only control type I error by Bradley Criteria when the sample sizes are small ($n = 5, 10, 15$) at α = 0.01 and when $n = 10$ at α = 0.05 and 0.10.

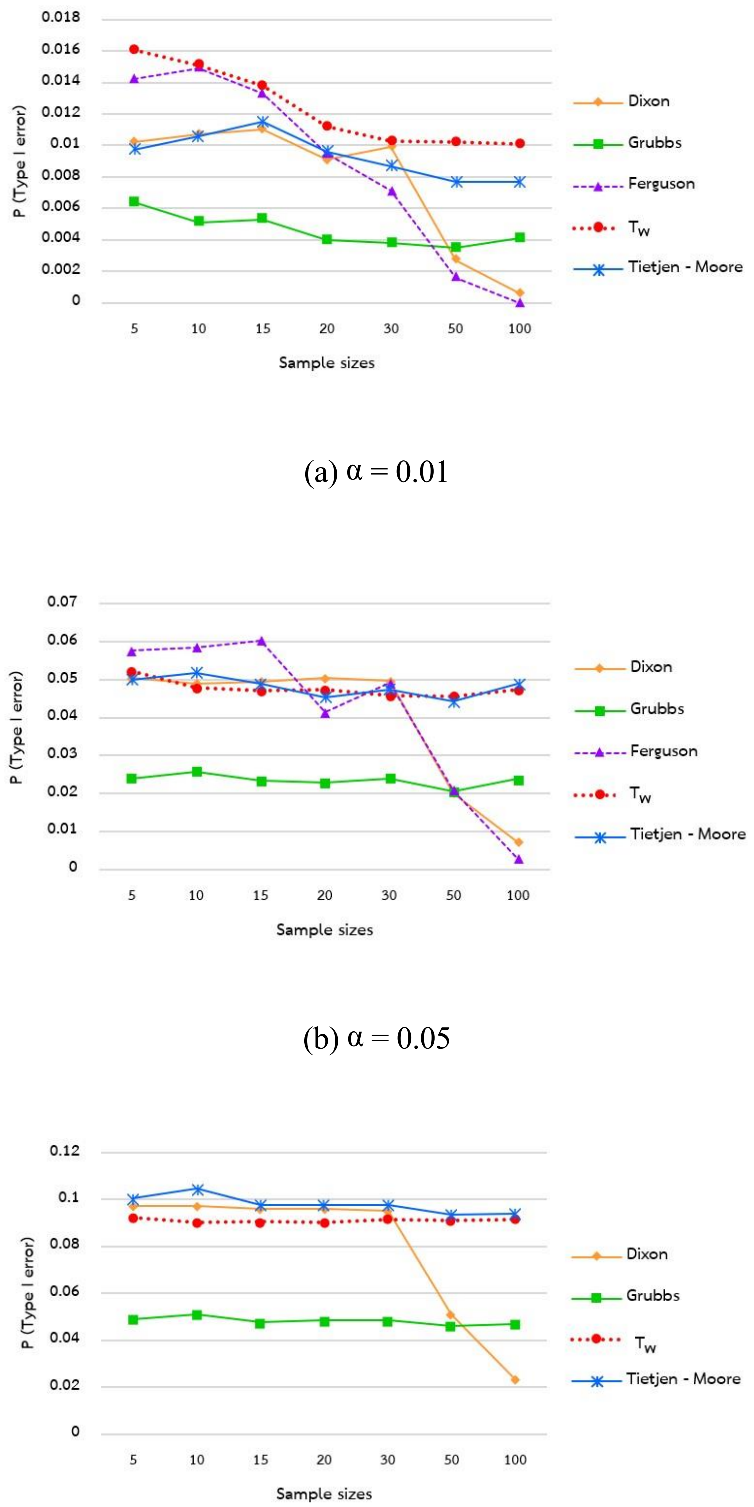
For our simulation study to compare the performance of Dixon's test, Ferguson's test, Grubbs' test, and T_W-test in terms of power of the test when there is one outlier in the sample, results of which are presented in Table 2 and Figure 3. From the detection of one mild outlier results, we found that T_W-test has the highest power when $n \leq 30$ compared to the other three tests (which are Dixon's test, Ferguson's test, and Grubbs' test) but it performs worse as the sample size increases. This makes Grubbs' test performs better than T_W-test when the sample size is large, say $n = 100$, since the power of Grubbs' test tends to increase as the sample size increases. If the outlier is extreme, T_W-test still has the highest power when $n \leq 30$ and performs better as the sample size increases as well as Dixon's test, Grubbs' test, and Ferguson's test at α = 0.01.

For our simulation study to compare the performance of Tietjen-Moore's Test and T_W-Test in terms of power when there are $k = 10\%n$ and $20\%n$ outliers in the sample, results of which are presented in Table 3 and Figures 4 and 5. From the results of the detection of k outliers, we can see that Tietjen-Moore's test performs better than T_W-test in every situation when there are $10\%n$ outliers in the sample either the outliers are mild or extreme but if there are $20\%n$ mild outliers in the sample, T_W-test has higher power than Tietjen-Moore's when $n \leq 30$.

TABLE 1. Probability of type I error of the statistical tests for detecting outliers

α	*n*	Dixon	Grubbs	Ferguson	TM	T_W
0.01	5	$0.0102^{*,**}$	0.0064^{*}	$0.0142^{*,**}$	$0.0097^{*,**}$	0.0161
	10	$0.0107^{*,**}$	0.0051^{*}	$0.0149^{*,**}$	$0.0106^{*,**}$	0.0151
	15	$0.0110^{*,**}$	0.0053^{*}	$0.0133^{*,**}$	$0.0115^{*,**}$	$0.0138^{*,**}$
	20	$0.0091^{*,**}$	0.0040	$0.0095^{*,**}$	$0.0096^{*,**}$	$0.0112^{*,**}$
	30	$0.0099^{*,**}$	0.0038	$0.0071^{*,**}$	$0.0087^{*,**}$	$0.0103^{*,**}$
	50	0.0027	0.0035	0.0016	$0.0077^{*,**}$	$0.0102^{*,**}$
	100	0.0006	0.0041	0.0000	$0.0077^{*,**}$	$0.0101^{*,**}$
0.05	5	$0.0504^{*,**}$	0.0240	$0.0578^{*,**}$	$0.0500^{*,**}$	$0.0523^{*,**}$
	10	$0.0490^{*,**}$	0.0258^{*}	$0.0585^{*,**}$	$0.0518^{*,**}$	$0.0479^{*,**}$
	15	$0.0495^{*,**}$	0.0233	0.0604^{*}	$0.0489^{*,**}$	$0.0471^{*,**}$
	20	$0.0506^{*,**}$	0.0228	$0.0414^{*,**}$	$0.0454^{*,**}$	$0.0473^{*,**}$
	30	$0.0497^{*,**}$	0.0239	$0.0492^{*,**}$	$0.0475^{*,**}$	$0.0458^{*,**}$
	50	0.0200	0.0206	0.0209	$0.0443^{*,**}$	$0.0455^{*,**}$
	100	0.0070	0.0238	0.0028	$0.0489^{*,**}$	$0.0475^{*,**}$
0.10	5	$0.0974^{*,**}$	0.0488		$0.1004^{*,**}$	$0.0924^{*,**}$
	10	$0.0974^{*,**}$	0.0509^{*}		$0.1049^{*,**}$	$0.0904^{*,**}$
	15	$0.0962^{*,**}$	0.0478		$0.0978^{*,**}$	$0.0906^{*,**}$
	20	$0.0962^{*,**}$	0.0484		$0.0979^{*,**}$	$0.0904^{*,**}$
	30	$0.0954^{*,**}$	0.0485		$0.0979^{*,**}$	$0.0916^{*,**}$
	50	0.0509^{*}	0.0460		$0.0936^{*,**}$	$0.0909^{*,**}$
	100	0.0233	0.0471		$0.0939^{*,**}$	$0.0916^{*,**}$

TM = Tietjen-Moore's test, * satisfy Bradley's criterion, ** satisfy Cochran's criterion



(a) α = 0.01

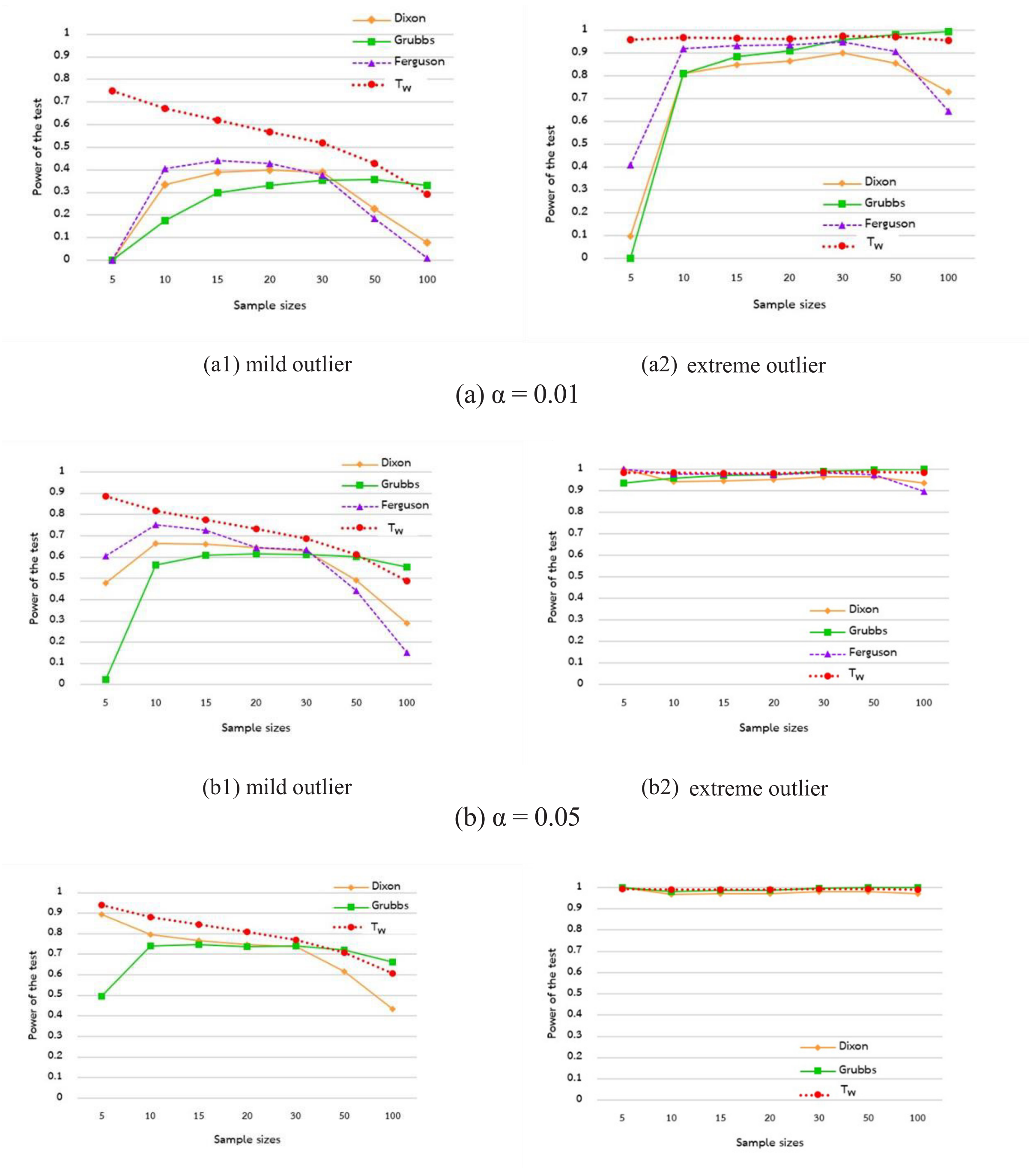
(b) α = 0.05

(c) α = 0.10

FIGURE 2. Probability of type I error of the statistical tests for detecting outliers against sample sizes *n* at significance levels (a) α = 0.01, (b) α = 0.05, and (c) α = 0.10

TABLE 2. Power of four statistical tests for detecting one outlier

		mild outlier				extreme outlier			
α	n	Dixon	Grubbs	Ferguson	T_W	Dixon	Grubbs	Ferguson	T_W
0.01	5	0.0000	0.0000	0.0000	0.7500	0.0994	0.0000	0.4121	0.9586
	10	0.3338	0.1755	0.4056	0.6718	0.8097	0.8112	0.9192	0.9670
	15	0.3908	0.2993	0.4404	0.6204	0.8475	0.8849	0.9315	0.9659
	20	0.3998	0.3301	0.4273	0.5692	0.8663	0.9112	0.9350	0.9621
	30	0.3885	0.3532	0.3755	0.5179	0.9009	0.9589	0.9497	0.9737
	50	0.2272	0.3567	0.1861	0.4278	0.8539	0.9808	0.9055	0.9720
	100	0.0797	0.3301	0.0109	0.2933	0.7308	0.9930	0.6444	0.9553
0.05	5	0.4796	0.0235	0.6049	0.8852	1.0000	0.9356	0.9999	0.9843
	10	0.6640	0.5640	0.7510	0.8173	0.9419	0.9570	0.9786	0.9839
	15	0.6599	0.6102	0.7267	0.7748	0.9454	0.9698	0.9774	0.9824
	20	0.6456	0.6151	0.6456	0.7321	0.9531	0.9750	0.9733	0.9825
	30	0.6275	0.6122	0.6352	0.6871	0.9663	0.9919	0.9838	0.9875
	50	0.4895	0.6011	0.4436	0.6116	0.9634	0.9971	0.9735	0.9881
	100	0.2878	0.5540	0.1505	0.4886	0.9352	0.9991	0.8966	0.9837
0.10	5	0.8929	0.4953		0.9397	1.0000	1.0000		0.9942
	10	0.7948	0.7419		0.8816	0.9683	0.9792		0.9898
	15	0.7656	0.7486		0.8442	0.9702	0.9872		0.9888
	20	0.7467	0.7369		0.8075	0.9719	0.9873		0.9893
	30	0.7378	0.7396		0.7705	0.9817	0.9969		0.9930
	50	0.6174	0.7213		0.7072	0.9813	0.9989		0.9927
	100	0.4332	0.6620		0.6053	0.9706	0.9999		0.9908



(a1) mild outlier

(a2) extreme outlier

(a) α = 0.01

(b1) mild outlier

(b2) extreme outlier

(b) α = 0.05

(c1) mild outlier

(c2) extreme outlier

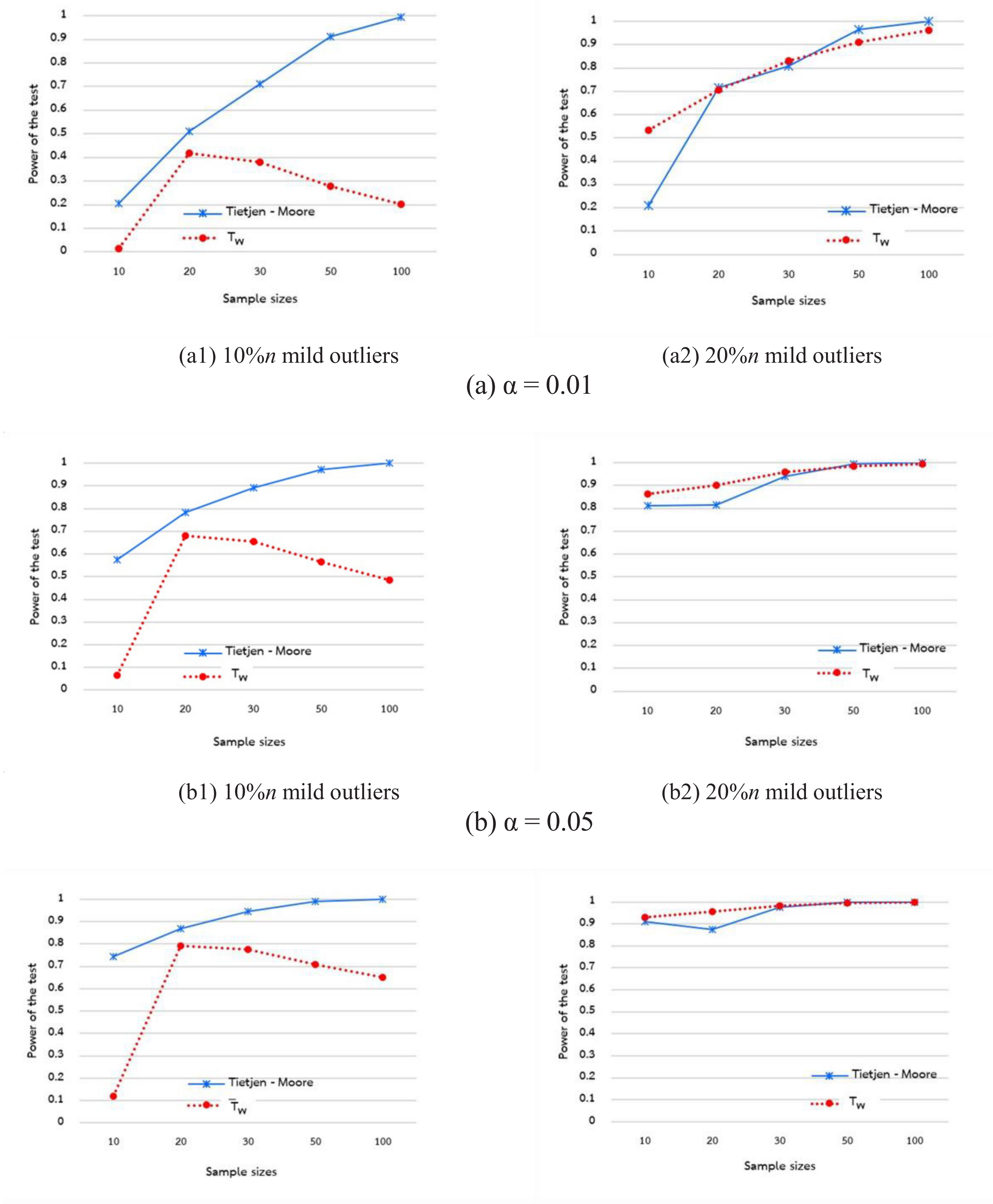
(c) α = 0.10

FIGURE 3. Graphs of power of four tests for detecting one outlier against sample sizes *n* at significant levels (a) α = 0.01, (b) α = 0.05, and (c) α = 0.10 when the outliers are mild and extreme

TABLE 3. Power of two statistical tests for detecting *k* outliers when *k* is 10% and 20% of sample sizes

		k mild outliers				*k* extreme outliers			
		10%*n*		20%*n*		10%*n*		20%*n*	
α	*n*	TM	T_W	TM	T_W	TM	T_W	TM	T_W
0.01	10	0.2036	0.0135	0.2087	0.5343	0.8332	0.0156	0.9263	0.9629
	20	0.5106	0.4174	0.7142	0.7061	0.9641	0.9475	0.9856	0.9888
	30	0.7095	0.3788	0.8093	0.8302	0.9961	0.9718	0.9967	0.9980
	50	0.9106	0.2790	0.9661	0.9106	1.0000	0.9812	1.0000	1.0000
	100	0.9951	0.2035	0.9996	0.9615	1.0000	0.9728	1.0000	0.9999
0.05	10	0.5736	0.0649	0.8121	0.8614	0.9585	0.0548	0.9918	0.9920
	20	0.7815	0.6814	0.8151	0.9009	0.9914	0.9813	0.9894	0.9978
	30	0.8928	0.6530	0.9382	0.9588	0.9994	0.9896	0.9997	1.0000
	50	0.9712	0.5647	0.9930	0.9845	1.0000	0.9935	1.0000	1.0000
	100	0.9997	0.4864	0.9999	0.9940	1.0000	0.9937	1.0000	0.9999
0.1	10	0.7437	0.1203	0.9106	0.9309	0.9793	0.1021	0.9955	0.9964
	20	0.8680	0.7903	0.8746	0.9574	0.9959	0.9895	0.9942	0.9993
	30	0.9458	0.7752	0.9755	0.9834	0.9997	0.9942	0.9998	1.0000
	50	0.9901	0.7101	0.9970	0.9953	1.0000	0.9965	1.0000	1.0000
	100	0.9999	0.6519	0.9999	0.9982	1.0000	0.9976	1.0000	1.0000

TM = Tietjen-Moore’s Test



(a1) 10%*n* mild outliers

(a2) 20%*n* mild outliers

(a) α = 0.01

(b1) 10%*n* mild outliers

(b2) 20%*n* mild outliers

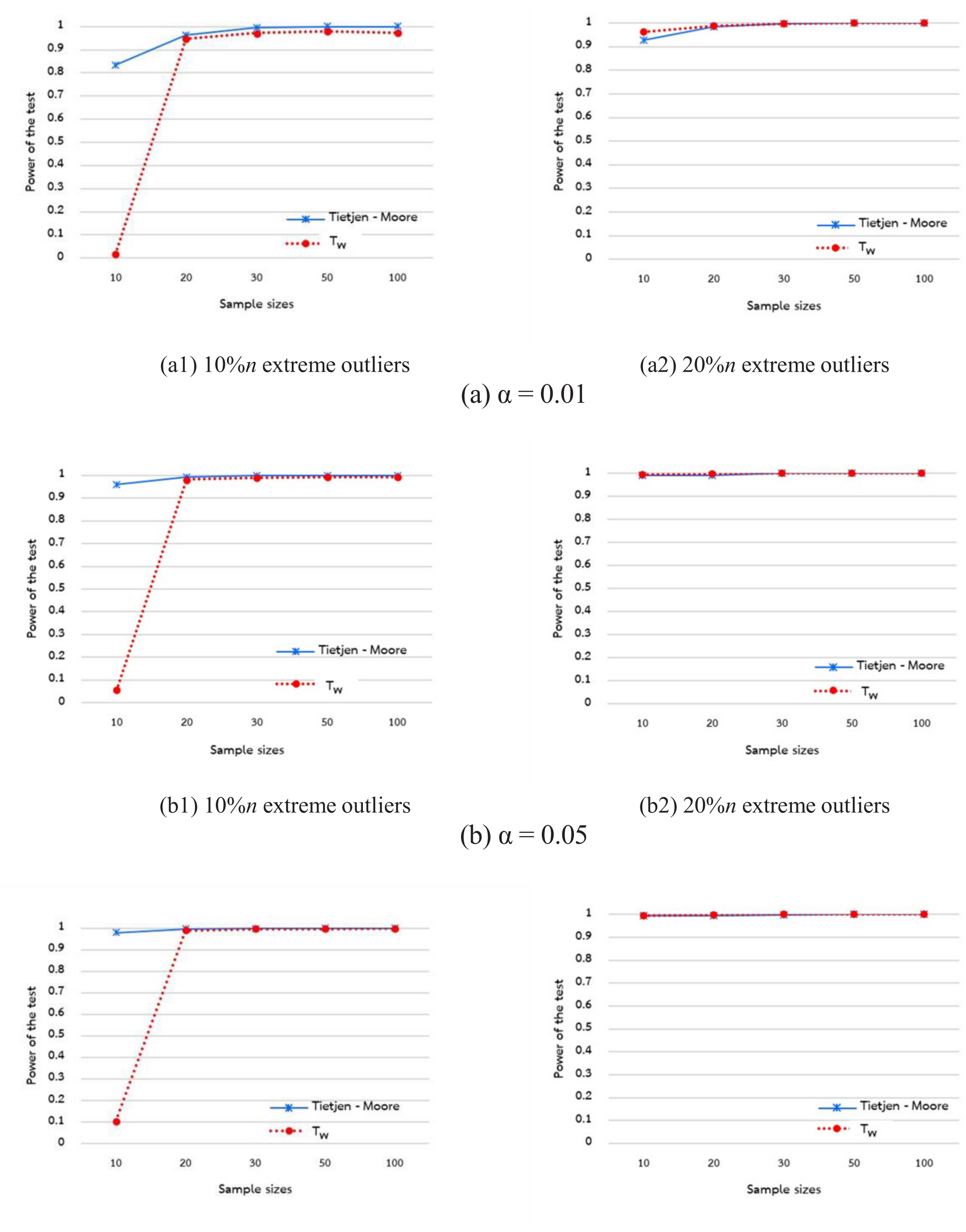
(b) α = 0.05

(c1) 10%*n* mild outliers

(c2) 20%*n* mild outliers

(c) α = 0.10

FIGURE 4. Graphs of power of two tests for detecting k = 10%n and k = 20%n outliers against sample sizes n at significant levels (a) α = 0.01, (b) α = 0.05, and (c) α = 0.10 (mild outliers)



(a1) 10%*n* extreme outliers

(a2) 20%*n* extreme outliers

(a) α = 0.01

(b1) 10%*n* extreme outliers

(b2) 20%*n* extreme outliers

(b) α = 0.05

(c1) 10%*n* extreme outliers

(c2) 20%*n* extreme outliers

(c) α = 0.10

FIGURE 5. Graphs of power of two tests for detecting k = 10%n and k = 20%n outliers against sample sizes n at significant levels (a) α = 0.01, (b) α = 0.05, and (c) α = 0.10 (extreme outliers)

TABLE 3. Power of two statistical tests for detecting k outliers when k is 10% and 20% of sample sizes

		k mild outliers				k extreme outliers			
		10%n		20%n		10%n		20%n	
α	n	TM	T_W	TM	T_W	TM	T_W	TM	T_W
0.01	10	0.2036	0.0135	0.2087	0.5343	0.8332	0.0156	0.9263	0.9629
	20	0.5106	0.4174	0.7142	0.7061	0.9641	0.9475	0.9856	0.9888
	30	0.7095	0.3788	0.8093	0.8302	0.9961	0.9718	0.9967	0.9980
	50	0.9106	0.2790	0.9661	0.9106	1.0000	0.9812	1.0000	1.0000
	100	0.9951	0.2035	0.9996	0.9615	1.0000	0.9728	1.0000	0.9999
0.05	10	0.5736	0.0649	0.8121	0.8614	0.9585	0.0548	0.9918	0.9920
	20	0.7815	0.6814	0.8151	0.9009	0.9914	0.9813	0.9894	0.9978
	30	0.8928	0.6530	0.9382	0.9588	0.9994	0.9896	0.9997	1.0000
	50	0.9712	0.5647	0.9930	0.9845	1.0000	0.9935	1.0000	1.0000
	100	0.9997	0.4864	0.9999	0.9940	1.0000	0.9937	1.0000	0.9999
0.1	10	0.7437	0.1203	0.9106	0.9309	0.9793	0.1021	0.9955	0.9964
	20	0.8680	0.7903	0.8746	0.9574	0.9959	0.9895	0.9942	0.9993
	30	0.9458	0.7752	0.9755	0.9834	0.9997	0.9942	0.9998	1.0000
	50	0.9901	0.7101	0.9970	0.9953	1.0000	0.9965	1.0000	1.0000
	100	0.9999	0.6519	0.9999	0.9982	1.0000	0.9976	1.0000	1.0000

TM = Tietjen-Moore's Test

CONCLUSION

Outlier detection is one of the most important tasks in applied research and outlier extraction is a problem in the process of discovering knowledge Discovery in Databases (KDD) which can improve the quality of the data and reduce the impact of having outliers in the sample. In this research, we studied the outlier detection methods in normal data by testing the hypothesis which is that the outliers come from whether the same or different populations. This work is on comparison of the efficiency of statistical tests for detecting outliers in terms of the probability of type I error and the power of the tests which can conclude that the T_W-test has the highest sensitivity in detecting one outlier when the sample size is small or moderate but, if the sample size is large, Grubbs' test performs better which is why many researchers have used and studied Grubbs' method (Grubbs 1969, 1950; Rahman, Sathik & Kannan 2014; Tietjen & Moore 1972). However, the performance of four test statistics which are Dixon's test, Ferguson's test, Grubbs' test, and T_W-test for detecting an extreme outlier is not that different especially when the significance level is 0.05 and 0.10. The power of Tietjen-Moore's test and T_W- test for detecting $k \geq 2$ outliers in the sample tends to increase as the sample size increases. Tietjen-Moore's test has higher sensitivity than T_W-test when k equals 10% of sample size, contrary to the case of when k makes up for 20%.

For future work, there are also other statistical tests for detecting outliers in the sample by using statistical hypothesis testing method which may have more efficiency. Also, some test statistics may be suitable for detecting outliers when the data has other distributions such as Weibull distribution, gamma distribution, or skew-normal distribution. Moreover, we recommend that the researchers may consider adjusting the variance of the sample to determine whether it affects the test power.

ACKNOWLEDGEMENTS

The authors would like to thank Faculty of Science, Prince of Songkla University, Hat Yai, Thailand since this work was supported by the Faculty of Science Research Fund, Prince of Songkla University.

REFERENCES

Barnett, V. & Lewis, T. 1984. *Outliers in Statistical Data*. New York: John Wiley & Sons.

Bradley, J.V. 1978. "Robustness?" *British Journal of Mathematical and Statistical Psychology* 31: 144-152.

Cochran, W.G. 1954. Some methods for strengthening the common χ^2 tests. *Biometrics* 10: 417-451.

Dixon, W.J. 1951. Ratios involving extreme values. *The Annals of Mathematical Statistics* 22: 68-78.

Dixon, W.J. 1953. Processing data for outliers. *Biometrics* 9: 74-89.

Efstathiou, C.E. 2006. Estimation of type I error probability from experimental Dixon's *q* parameter on testing for outliers within small size data sets. *Talanta* 69: 1068-1071.

Ferguson, T.S. 1961. On the rejection of outliers. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*. Berkeley: University of California Press. pp. 253-287.

Grubbs, F.E. 1950. Sample criteria for testing outlying observations. *The Annals of Mathematical Statistics* 21: 27-58.

Grubbs, F.E. 1969. Procedures for detecting outlying observations in samples. *Technometrics* 11: 1-21.

Hawkins, D.M. 1980. *Identification of Outliers*. London: Chapman and Hall.

Jareankam, W. 2020. A detection of outliers in random sample from normally distributed population using coefficient of skewness. *Burapha Science Journal* 25: 236-245.

Jareankam, W. 2013. A detection of outliers in random sample from normal population. Thesis. National Institute of Development Administration (Unpublished).

Patchayaluck, N. 2013. Estimation of probability of type I error and power of test for test statistics an outlier. Thesis, Silpakorn University (Unpublished).

Rahman, S.K., Sathik, M.M. & Kannan, K.S. 2014. A novel approach for univariate outlier detection. *International Journal of Scientific & Engineering Research* 5: 1594-1599.

Rattanaloetnusorn, S. 1991. A comparative study on some procedures for detecting outliers in linear regression analysis. Thesis, Chulalongkorn University (Unpublished).

Rosner, B. 1975. On the detection of many outliers. *Technometrics* 17: 221-227.

Tietjen, G.L. & Moore, R.H. 1972. Some grubbs-type statistics for the detection of several outliers. *Technometrics* 14: 583-597.

Verma, S.P. & Quiroz-Ruiz, A. 2006. Critical values for six Dixon tests for outliers in normal samples up to sizes 100, and applications in science and engineering. *Revistamexicana de Ciencias Geológicas* 23: 133-161.

*Corresponding author; email: phontita.t@psu.ac.th