Published by
NIGERIAN SOCIETY OF PHYSICAL SCIENCES
Available online @ https://asr.nsps.org.ng/index.php/asr

African Scientific Reports 1 (2022) 142–153

African Scientific Reports

A Simulation Study on Robustness of One Sample Inferential Statistics in Mixture Distribution

T. J. Adejumo[a,*], A. A. Akomolafe[b], A. I. Okegbade[a], S. D. Gbolagade[a]

[a]*Department of Statistics, Ladoke Akintola University of Technology, Ogbomoso, Oyo State, Nigeria*
[b]*Department of Statistics, Federal University of Technology, Akure, Ondo State, Nigeria*

Abstract

Mixture distribution refers to the combination of more than one probability distribution. Meanwhile, non-normality of data set may be inevitable and the cause may be as a result of mixed distributions thereby renders parametric tests ineffective. Montecarlo experiment was performed 5000 times under twelve sample sizes where data were generated from Gaussian and Cauchy distributions using R-statistical packages. At three commonly used alpha levels (0.1, 0.05 and 0.01), the robustness of the test statistics (Rank transformation t-test, Wilcoxon sign test (Distribution and Asymptotic), Signed rank test (Distribution and Asymptotic) and Trimmed t-test) were examined. When the type I error rate of a statistic approximately equal to the true error rate then the statistic is considered robust. At 0.1 and 0.05, Rank transformation t-test, Wilcoxon sign test (distribution) and Trimmed t-test in this order are robust. Meanwhile, at 0.01 Rank transformation and Wilcoxon sign test (distribution) were identified to be robust. Also, further counts at all levels of significance revealed that the Rank transformation test is robust and thereby recommended when data comes from a mixed distribution. Hence, this study has been able to identify test statistics that are robust when data comes from a mixed distribution in one sample problem.

DOI:10.46481/asr.2022.1.3.35

Keywords: Inferential statistics, Mixture distribution, Non-normality, Simulation study

Article History :
Received: 23 May 2022
Received in revised form: 22 August 2022
Accepted for publication: 03 November 2022
Published: 29 December 2022

Communicated by: Tolulope Latunde

1. Introduction

A probability distribution that is derived from the combination of two or more random variables at different distributions is called mixture distribution. In other words, a mixture distribution occurs when more than one probability

*Corresponding author tel. no: +234 8066160296
Email address: tjadejumo@lautech.edu.ng (T. J. Adejumo)

distributions are combined to form a single distribution. Data when analyzed often fail the assumption of normality which could be as a result of unequal variances in the error terms or presence of extreme observations in the data set, and thus the need for equivalent non parametric tests. When data does not come from a known distribution, then the random variables are not identical, such data usually come from a mixed distribution. It also exists as finite or countable mixtures and it at times exists as uncountable mixtures.

There are several literature on mixture distribution in diverse areas of specialization. Such include the social and behavioral sciences, environmental sciences, engineering and physical sciences etc [1, 2-4]. In biological and physical sciences most especially, [5] illustrated that mixture of distributions do occur such that if random sample of fish species is taken, therefore the characteristics measured for each member of the sample will definitely vary with age but, the distribution of the characteristics in all population will be a mixture of the distributions at different ages. [6] considered mixture models when the mixing distribution can be quietly identified using Schwarz's criteria and Neyman test. In his analysis, he presented smooth goodness of fits for testing the mixture distribution of a sequence of independently identically distributed random variables.

In case of [7], using the likelihood ratio (LR) test for unconditional geometric distributions examined the mixture hypothesis of conditional geometric distributions. Through simulation studies, the interrelationship between geometric and exponential mixture hypothesis was examined. Meanwhile, [8] claimed that under a Dirichlet process prior unobserved random effects contribute to unequal variance of the error terms among sampling units and therefore, smooth nonparametric estimate of mixture distribution can be derived as an approximate nonparametric Bayes estimate. Also, in [9] with the aids of Monte Carlo experiment, the relative power of paired parametric and nonparametric tests were assessed. The outcome of their results revealed that, in given situation each statistic was more powerful. Following the proposed distribution of [10] through simulation studies studied the effect of the Gaussian distribution assumptions on the size of one sample nonparametric test (Sign test) and its power where one sample ttest and Wilcoxon's signed rank test were employed in one sample location problem. Under various Kurtosis and Skewness criterions, the power functions of the test statistics considered were generated for several sample sizes. One of their results shows that all tests are more powerful for lighter tails rather heavier tails, with the effect strongest for the sign test, followed by Wilcoxon test and then t-test.

According to [11], for the one and two samples problem at different degrees of correlations, sample sizes, population distributions including true difference in location by analysis investigated the behavior of the student t-test and the Wilcoxon test using Monte-Carlo simulations. Their results revealed that the two sample tests performed below expectation in terms of type I error rates performance. Sequel to the outcome of their results, they affirmed that the sign of the underlying correlation cannot be used as a basis to conclude that which of the test is to be used between the t - test or the Wilcoxon test for one sample tests. Presence of extreme observation in the data set may be inevitable even in paired observations, this made [12] to examine the performance of some paired inferential statistics in the presence of outliers where Paired t-test, Wilcoxon sign rank test, Rank transformation t - test and Trimmed t-test were considered as inferential statistics. Through simulation studies, data were obtained from Gaussian distribution and polluted with degrees of outliers and multicollinearities. Under different levels of multicollinearities and alpha levels, they concluded that Rank transformation test, Distribution Sign test and Trimmed t-test statistics respectively can accommodate outliers.

In the literature, authors have examined the robustness of some inferential statistics in the presence of outliers in one and paired samples problem at different levels of multicollinearity and significance levels when data are only generated from normal distribution whereas other distributions were not put into consideration. Hence, to bridge this gap, this study examines the robustness of some one-sample inferential test statistics when data comes from mixture distribution. The distribution considered in the study where data was generated are normal and Cauchy distributions. Without any loss of generality, this study in the long run was able to identify some non-parametric and semi-parametric inferential test statistics that are robust when data comes mixture distribution at different sample sizes and levels of significance. The distributions and the simulation procedures are discussed as follows.

2. Materials and methods

2.1. *Distribution used for the Study*

In this study, data was generated from two distributions. These include the normal distribution and the Cauchy distribution. Both are briefly illustrated below

2.1.1. Normal distribution:

One of the most commonly used distributions is the Gaussian distribution otherwise known as the normal distribution; this is because of its ability to approximate several natural phenomena. It serves as reference for many probability distributions. It is a distribution that explains how the values of a variable are distributed. In normal distribution, most of the observations are cluster around the mean.

2.1.2. Properties of the Normal distribution

1. It is symmetric about the mean and has bell shaped
2. Its random variable ranges from -∞ to ∞
3. It has two parameters, μ and σ.

The normal density function is

$$f\left(x : \mu, \sigma^2\right) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2} \tag{1}$$

This study employed an hypothesized mean ($\mu = 10$) and variance ($\sigma^2 = 25$).

2.1.3. Cauchy distribution

A continuous probability distribution that has the ratio of two independently normally distributed random variables with the mean at the denominator of the distribution zero is called Cauchy distribution. It is otherwise known as the Lorentz distribution. It is a distribution that its mean and variance are undefined hence, it is called a "pathological" distribution. Its peak are tall other than normal distribution whereas its fat tails slowly decayed gradually.

The probability density function of Cauchy distribution is defined as follows;

$$pdf = \frac{1}{\pi\gamma\left[1 + \left(\frac{x - x_0}{\gamma}\right)^2\right]} \tag{2}$$

where γ is the scale parameter.

The special case of the student t distribution is the standard Cauchy distribution with the degrees of freedom $n - 1 = 1$.

i.e. t (df = 1):

$$pdf = \frac{\Gamma\left(\frac{\upsilon+1}{2}\right)}{\Gamma\left(\frac{\upsilon}{2}\right)\sqrt{\upsilon\pi}}\left(1 + \frac{x^2}{\upsilon}\right)^{-\frac{\upsilon+1}{2}} = \frac{1}{\pi\left(1 + x^2\right)} \tag{3}$$

$$where\ \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$$

2.1.3.1 Properties of Cauchy distribution:

Since the pdf of Cauchy distribution is an even function about $x = a$ (i.e the median) therefore pdf is symmetric about the line $x = a$.

1. The pdf has undefined mean.
2. The pdf has undefined variance

2.2. Review of some inferential Statistics

2.2.1. Trimmed t-test (One sample)

In order to combat the effect of extreme values in the data set, [13] came up with another inferential statistics in this regard called trimmed t –test. In trimmed mean, the average is adjusted by reducing the entire data by certain percentage at the upper and lower tail of the data. After reducing the specified extreme values, it is then found using a standard arithmetic averaging formula. Also, he proposed another test statistic for the two samples problem, when the population variances are not equal that excludes extreme observations [14]. In this case, q – observations were removed from each tail of the distribution. Equation 5 defined the statistic.

$$\overline{X_t} = \frac{X_{q+1} + X_{g+2} + \cdots + X_{n-q}}{n-2q} \tag{4}$$

Such that $x_1,...,x_n$ are the ordered observations in the data, q are the values that have been reduced at the beginning of the sample data. Likewise, the denominator of equation (4) indicates the number of data deleted from the sampled data.

Given the Winsorized mean, its sum-of-squared derivation can be computed as:

$$\mathrm{SSD_w} = [q+1]\left[x_{q+1} - \overline{X}_w\right]^2 + \left[x_{q+2} - \overline{X}_w\right]^2 + \ldots + [q+1]\left[x_{n-q} - \overline{X}_w\right]^2 \tag{5}$$

And its variance can be expressed as:

$$S_w^2 = \frac{SSD_w}{n-2q-1} \tag{6}$$

The one sample test for trimmed test can be defined as:

$$t_w = \frac{\overline{X}_t - \mu_t}{S_w \big/ \sqrt{n-2q-1}} \tag{7}$$

where$t_w$is the test statistic for the trimmed test and it follows a t-distribution and $\overline{X}_t$is the mean of the remaining observation after q-observation has been removed.

2.2.2. *The sign test*

Sign test is a one sample nonparametric test discovered by [15] in early 17th century. Its main advantage is to test whether or not two groups are equally sized especially when data comes from unknown population and when sampled data are ordered in pairs. The numerical magnitudes of the observations are not based but rather on the signs attached to the values, which is either plus (+) or minus (-). With probability value 0.5, it is otherwise known as binomial sign test. In addition, it is used to test the pair value below or above the median and the pair difference is not measured hence, it is considered a weak test. The test is rejected if

$P(W \leq W^+) < \alpha \text{ or } P(W \leq W^-) < \propto$, when the test is two-tail, $\propto/2$ is used instead of α.

Asymptotically, the test has the distribution of binomial (n, 0.5) as in: ${}^{n}C_{x}\, q^{n-x}\, p^{x}$ with $p_x = 0.5$. Its asymptotic test statistic is:

$$Z = \frac{W^+ - \frac{n}{2}}{\sqrt{\frac{n}{4}}} \sim N(0,1) \tag{8}$$

2.2.3. *Rank Transformation*

In order to circumvent the bridge between parametric and nonparametric tests [16] proposed Rank transformation test. Here, instead of the data itself, the usual parametric procedures are applied to the ranks of the data. It was seen as an alternative method just as in nonparametric methods to solve problems if such data does not follow a normal distribution.

One sample T-test for Rank Transformation

According to [16], the test statistic was defined as follows:

Let $D_1, D_2, \ldots, D_n$ indicate independent random variables with a common mean.

Meanwhile, for matched pairs (X_i, Y_i);$D_i = X_i, - Y_i$. The statistic is given as:

$$T = \frac{\sum_{i=1}^{n} R_i}{\sqrt{\sum_{i=1}^{n} {R_I}^2}} \tag{9}$$

Where R_i = (sign D_i) x (rank of / D_i/)

An equivalent formula to (9) is (10)

$$t_R = \frac{\sum_{i=1}^{n} R_i}{\sqrt{\frac{n\sum_{i=1}^{n} {R_I}^2 - \left(\sum_{i=1}^{n} R_i\right)^2}{n-1}}} \tag{10}$$

Equation (10) has the t-distribution with (n-1) degree freedom.

However, (10) can further be expressed as:

$$t_R = \frac{T}{\sqrt{\frac{n}{n-1} - \frac{T^2}{n-1}}} \tag{11}$$

Which is defined to be a monotonic function of T.

2.3. Algorithm for simulation

How data were generated from different distributions and subjected to the inferential test statistics including the estimation of Type I error rates using Monte Carlo procedures with the aid of R-programming codes are hereby discussed.

2.3.1. Source of Data

The data used in this study were generated with the aid of R-statistical programming package using the following parameters

1. (n) = 10, 20, 30, 40, 60, 80, 100,200, 300, 400, 500 and 600 as sample sizes
2. Hypothesized median (md) and mean = 10
3. Standard deviation (δ) = 5
4. Replications (RR) = 5000
5. Presented α-level = 0.1, 0.05 and 0.01

2.3.2. Distributions used with the parameters.

The data were generated from the following distributions

1. Normal distribution with (μ) = 10 and (δ) = 5
2. Cauchy distribution (n, 10)

2.3.3. The Test Statistics used.

The test statistics used in the one sample problem are as follows:

1. One sample t-test for Rank transformation by [16]
2. Wilcoxon sign Test (Distribution (WSTD) and Asymptotic (WSTA)) by [17]
3. Wilcoxon signed Rank test (Distribution (WSTD) and Asymptotic (WSTA)) by [17]
4. Trimmed t-test (Tt-test) by [13]

2.3.4. Procedures for Data Generation and Estimation of Type I error Rate.

The procedures for data generation and calculation of Type I error rate in one sample mixture distribution are as follows:

1. Choose a sample size(n)
2. Generate random sample size from the distributions under consideration, for example X~N(n, 10, 5) and Cauchy~(n, 10).
3. Combine the data generated in step (ii).
4. Subject each considered statistic and save their p-values
5. Assign step(IV) as follows;

$$H_i = \begin{cases} 1, \; if \; p-value < \propto \\ 0, \; otherwise \end{cases}$$

where α = 0.1, 0.05 and 0.01 are the level of significance

1. From step (II) to (V) repeat up to 5000 times, RR=5000
2. At each statistics, add the results obtained in (VI) as in the equation below;

$$H = \sum_{i=1}^{RR} H_i$$

3. To estimate the type I error of the statistics, divide the result in (VII) by the (RR) as given as follows:

i.e.$K_\propto = \frac{\sum_{i=1}^{RR} H_i}{RR} = \frac{H}{RR}$
ix. Select (n) and follow the procedures from (II) - (VIII) until all (n) are completely used.

2.3.5. Examination of the test statistics

Robustness of the inferential statistics were investigated in mixture distribution. Any calculated Type 1 error rates of the test that falls within the range of 0.095 – 0.14, 0.045 – 0.054 and 0.005 – 0.014 for 0.1, 0.05 and 0.01 respectively at different alpha level (α)and sample sizes (n) which was adopted by [18], used by [3] and [1]. Also, a test statistic that has the highest number of counts is considered robust.

3. Results and Discussion

Here, the results of simulation for all the inferential statistics in mixture distribution of one sample problem including graphical representation are discussed.

3.1. Type I Error rate of One Sample Test Statistics in Mixture Distribution when α = 0.1:

Table 1 and 2 display the simulation results of the test statistics at 0.1 level of significance and graphically represented in figure 1a and 1b. RK-ttest, WSTD, Tt-Test, WSTA, WSRTD have better Type I error rate as the sample size increases while WSRTA are not good. Meanwhile when counted over the sample sizes, it was observed that RK-ttest is robust, followed by WSTD and Tt-test in that order.

3.2. Type I Error rate of One Sample Test Statistics in Mixture Distribution when α = 0.05:

Results of Monte Carlo experiments for each inferential statistics is presented table 3, 4 and graphically represented in figure 3, 4. The results and graphical representation indicate that the Type I error rate of RK-ttest, WSTD, Tt-Test, WSTA and WSRTD are good as the sample size increases while, WSRTA is not good. At the same time when counted over the sample sizes, it was observed that TK-ttest, WSTD and Tt-test in that order performed better while WSRTD, WSTA and WSRTA do not.

Table 1. Simulation Results at 0.1 Level of Significance

n	RK-ttest	WSTD	WSTA	WSRTD	WSRTA	Tt-test
10	**0.1124**	0.0904	0.0246	0.0246	0.6148	0.0582
20	**0.0954**	**0.0954**	0.038	0.038	0.8688	0.0746
30	**0.097**	0.0928	0.0928	0.0928	0.9526	0.0772
40	**0.0958**	0.094	0.0846	0.0846	0.9808	0.0882
60	0.0942	0.0942	0.0942	**0.0948**	0.998	0.0862
80	**0.0996**	**0.0996**	0.0942	0.0942	0.9996	**0.097**
100	**0.103**	**0.103**	0.0894	0.0894	1	**0.0982**
200	**0.0976**	**0.0976**	0.0726	0.0728	1	0.0944
300	**0.1014**	**0.1016**	**0.1024**	**0.1022**	1	**0.1008**
400	**0.1084**	**0.1084**	**0.1002**	**0.1002**	1	**0.1052**
500	**0.1068**	**0.1068**	**0.1048**	**0.1048**	1	**0.1014**
600	**0.0972**	**0.0972**	0.0944	0.0946	1	**0.101**

Source: Simulation Results

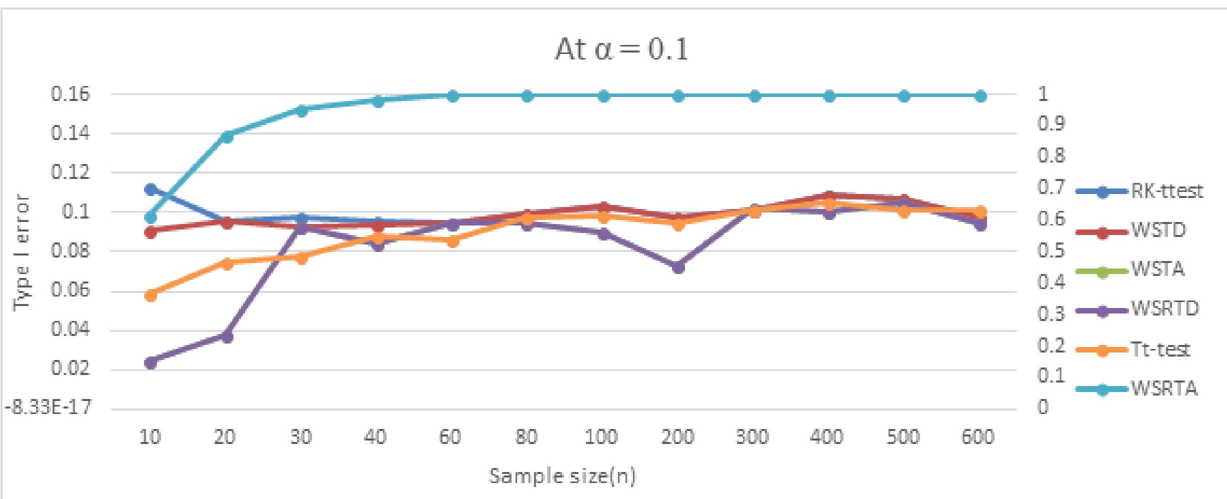

Figure 1. Graphical Representation of Type I Error rate of One Sample Test Statistics in Mixture Distribution when α = 0.1

Table 2. Times Type I error rate approximates to α = 0.1.

Alpha Level	Test statistics	10	20	30	40	60	80	100	200	300	400	500	600	SUM	RANK
	RK-ttest	1	1	1	1	0	1	1	1	1	1	1	1	11	1
	WSTD	0	1	0	0	0	1	1	1	1	1	1	1	8	2
α = 0.1	WSTA	0	0	0	0	0	0	0	0	1	1	1	0	3	5
	WSRTD	0	0	0	0	1	0	0	0	1	1	1	0	4	4
	WSRTA	0	0	0	0	0	0	0	0	0	0	0	0	0	6
	Tt-test	0	0	0	0	0	1	1	0	1	1	1	1	6	3

Source: Counted from Table 1

3.3. Type I Error rate of One Sample Test Statistics in Mixture Distribution when α = 0.01

Results of Monte Carlo experiments for each inferential statistics is presented table 5, 6 and graphically represented in figure 5, 6. From the results and graphical representation it can observed that the Type I error rate of RK-ttest, WSTD, Tt-Test, WSTA, WSRTD and are good as the sample size increases whereas, WSRTA weak. When

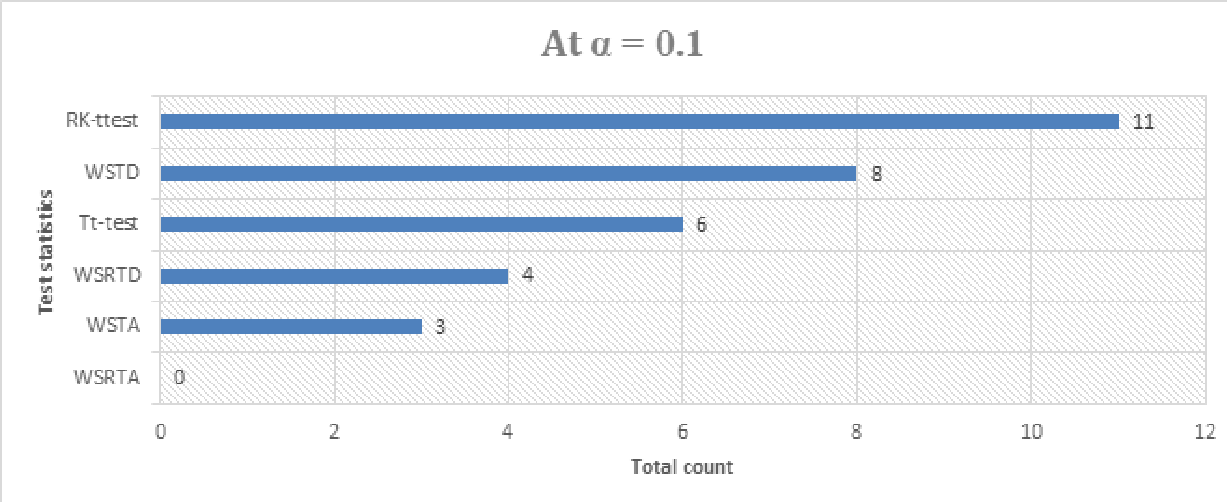

Figure 2. Bar chart indicating total times Type I error rates approximates to α = 0.1

Table 3. Simulation Results at 0.05 Level of Significance

n	RK-ttest	WSTD	WSTA	WSRTD	WSRTA	Tt-test
10	**0.0516**	**0.0516**	0.0246	0.0246	0.6148	0.0198
20	**0.0506**	**0.0464**	0.038	0.038	0.7476	0.0318
30	**0.048**	**0.048**	0.0386	0.0386	0.8988	0.0346
40	**0.0538**	**0.0528**	0.0386	0.0386	0.9584	0.0432
60	**0.046**	**0.045**	**0.045**	0.0228	0.9954	0.0392
80	**0.0498**	**0.0488**	0.034	0.034	0.9986	**0.0456**
100	**0.0512**	**0.0508**	0.0312	0.0312	0.9998	0.0434
200	**0.0534**	**0.0532**	0.0416	0.0418	1	**0.0494**
300	**0.0508**	**0.0506**	**0.0468**	**0.0466**	1	**0.0508**
400	**0.0516**	**0.0514**	0.0408	0.0408	1	**0.0484**
500	**0.0582**	**0.058**	**0.0484**	**0.0484**	1	**0.0542**
600	**0.0484**	**0.0484**	**0.0474**	**0.0474**	1	**0.0472**

Source: Simulation Results

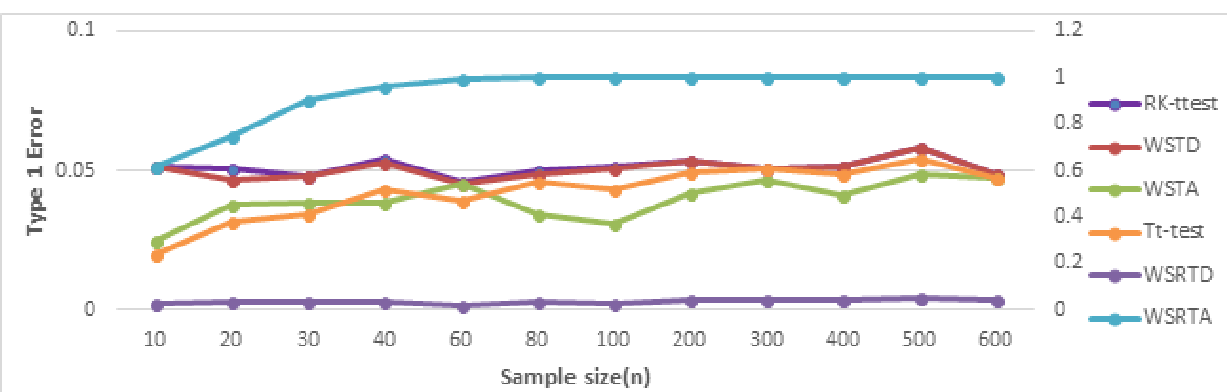

Figure 3. Graphical Representation of Type I Error rate of One Sample Test Statistics in Mixture Distribution when α = 0.05

!t

Table 4. Times Type I error rate approximates to α = 0.05

Alpha Level	Test statistics	10	20	30	40	60	80	100	200	300	400	500	600	SUM	RANK
	RK-ttest	1	1	1	1	1	1	1	1	1	1	1	1	12	1.5
	WSTD	1	1	1	1	1	1	1	1	1	1	1	1	12	1.5
α = 0.05	WSTA	0	0	0	0	1	0	0	0	1	0	1	1	4	4
	WSRTD	0	0	0	0	0	0	0	0	1	0	1	1	3	5
	WSRTA	0	0	0	0	0	0	0	0	0	0	0	0	0	6
	Tt-test	0	0	0	0	0	1	0	1	1	1	1	1	6	3

Source: Counted from Table 3

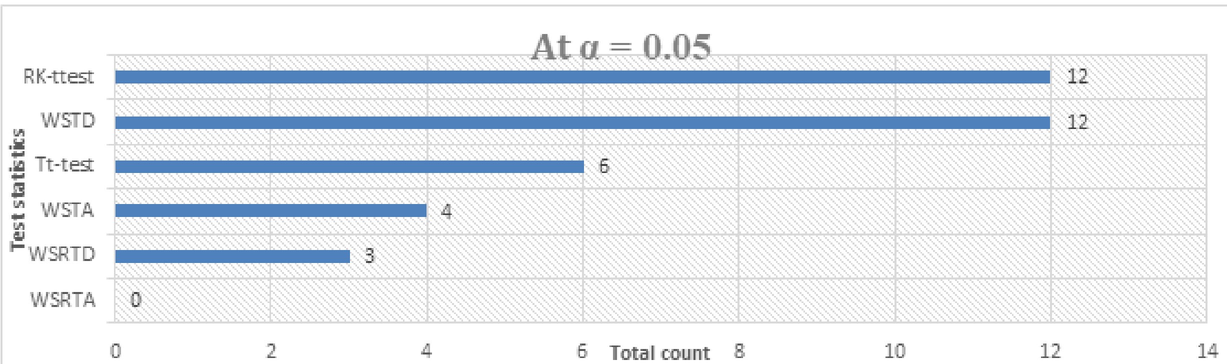

Figure 4. Bar chart indicating total times Type I error rates approximates to α = 0.05

Table 5. Simulation Result at 0.01 Level of Significance

n	RK-ttest	WSTD	WSTA	WSRTD	WSRTA	Tt-test
10	0.0164	**0.011**	0.0018	0.0018	0.3818	0.0024
20	**0.0118**	**0.0096**	0.0024	0.0024	0.5912	0.0036
30	**0.0096**	**0.0092**	**0.005**	**0.005**	0.8204	0.0046
40	**0.0108**	**0.0094**	**0.0052**	**0.0052**	0.9214	**0.0056**
60	**0.0082**	**0.0076**	**0.0076**	**0.0052**	0.9748	0.0046
80	**0.0118**	**0.0108**	**0.0088**	**0.0088**	0.9956	**0.0092**
100	**0.0104**	**0.0098**	**0.0078**	**0.0078**	0.9992	**0.0088**
200	**0.0094**	**0.0094**	**0.009**	**0.009**	1	**0.011**
300	**0.0108**	**0.0106**	**0.0114**	**0.0114**	1	**0.0104**
400	**0.011**	**0.011**	**0.009**	**0.009**	1	**0.0088**
500	**0.0102**	**0.0096**	**0.0092**	**0.0092**	1	**0.0102**
600	**0.0088**	**0.0084**	**0.0078**	**0.0078**	1	**0.0078**

Source: Simulation Results

counted over the sample sizes, it was observed that the statistics did well which includes; TK-ttest, WSTD, WSRTD, WSTA and Tt-test.

3.4. Result of Type I Error Rates for one sample when counted over all levels of significance.

The results in Table 7 and Figure 7 show that the RK-ttest and WSTD are robust in mixture distribution as they have highest count across all levels of significance.

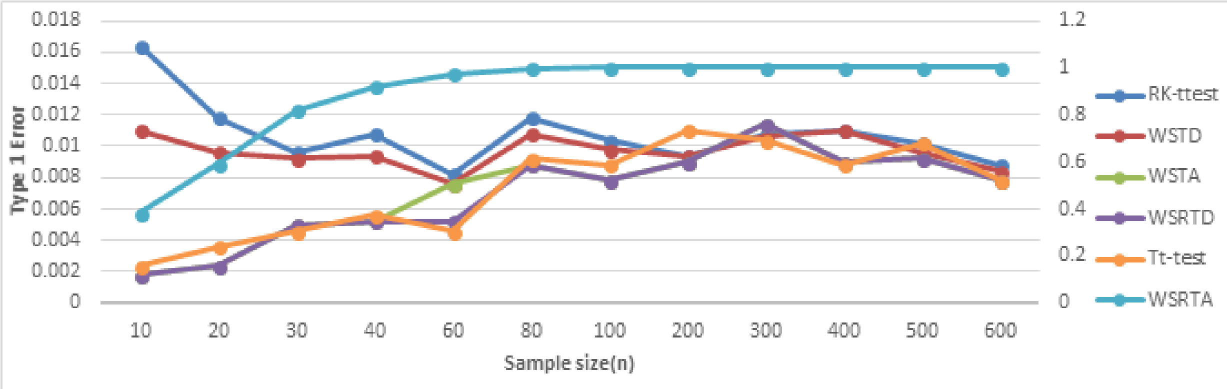

Figure 5. Graphical Representation of Type I Error rate of One Sample Test Statistics in Mixture Distribution when α = 0.01

Table 6. Times Type I error rate approximates to = 0.01

Alpha Level	Test statistics	10	20	30	40	60	80	100	200	300	400	500	600	SUM	RANK
	RK-ttest	2	3	3	3	2	3	3	3	3	3	3	3	**34**	1
	WSTD	2	3	2	2	2	3	3	3	3	3	3	3	32	2
α = 0.01	WSTA	0	0	1	1	2	1	1	1	3	2	3	2	17	4.5
	WSRTD	0	0	1	1	2	1	1	1	3	2	3	2	17	4.5
	WSRTA	0	0	0	0	0	0	0	0	0	0	0	0	0	6
	Tt-test	0	0	0	1	0	3	2	2	3	3	3	3	20	3

Source: Counted from Table 5

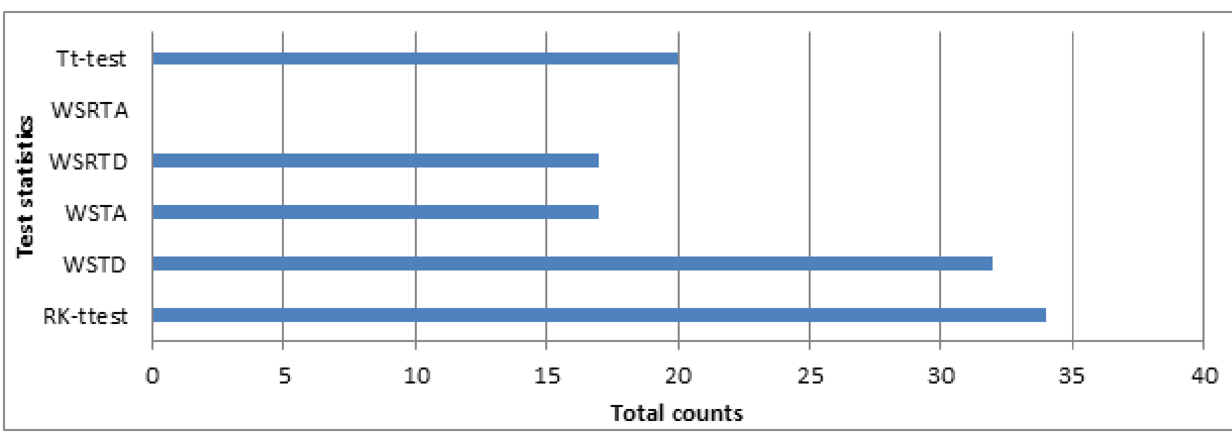

Figure 6. Bar chart indicating total times Type I error rates approximates to α = 0.01

4. Discussion

Following the investigation of simulation results of one sample inferential statistics at different levels of significance in which they have been presented in Table 1, 3 and 5 and as well represented graphically in Figure 1, 3 and 5 as the sample size increases. Likewise, the number of times the Type I error rates of the test statistics falls within the preferred interval is counted over the sample sizes. Hence, table 8 hereby summarizes the investigation for robustness of one sample inferential statistics in mixture distribution in order of importance.

Table 7. Total number Times Type I error rate approximates to true error rates when counted across the sample sizes

Test statistics	10	20	30	40	60	80	100	200	300	400	500	600	SUM	RANK
RK-ttest	2	3	3	3	2	3	3	3	3	3	3	3	**34**	**1**
WSTD	2	3	2	2	2	3	3	3	3	3	3	3	32	2
WSTA	0	0	1	1	2	1	1	1	3	2	3	2	17	4.5
WSRTD	0	0	1	1	2	1	1	1	3	2	3	2	17	4.5
WSRTA	0	0	0	0	0	0	0	0	0	0	0	0	0	6
Tt-test	0	0	0	1	0	3	2	2	3	3	3	3	20	3

Source: Counted from Table 2, 4 and 6

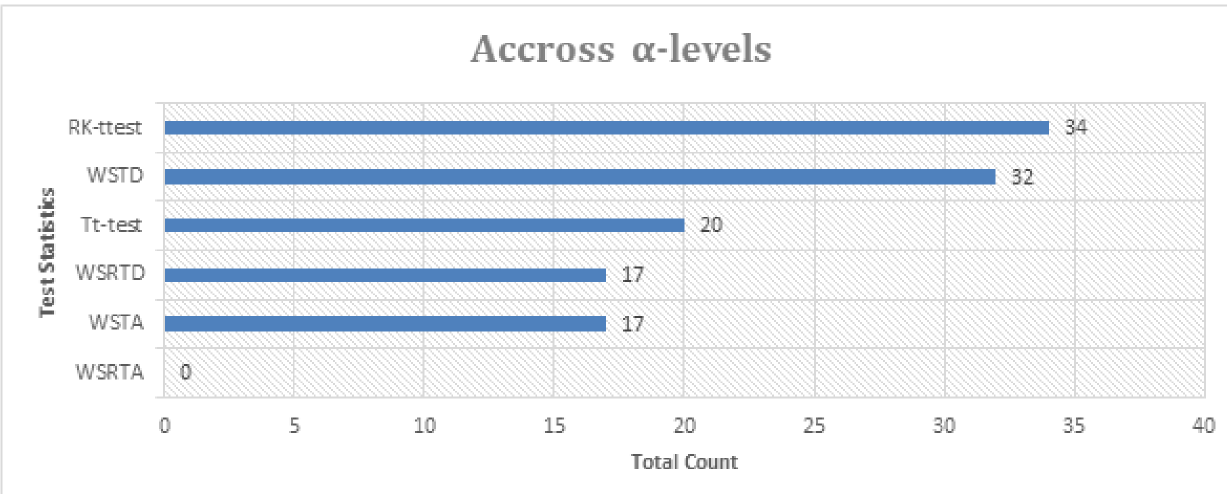

Figure 7. Bar chart indicating overall total times Type I error rates approximates to the true error rates across all sample sizes

Table 8. Overall Summary of Robustness of One sample inferential Statistics in Mixture Distribution

Alpha level	**One Sample Investigation**
	Robustness
0.1	Rk-ttest, WSTD and Tt-test
0.05	Rk-ttest, WSTD and Tt-test
0.01	Rk-ttest, WSTD, Tt-test and WSTA
Overall	Rk-ttest, WSTD and Tt-test

5. Conclusion

When $\alpha = 0.1$ RK-ttest, WSTD and Tt-test in that order are more robust in a mixture distribution and $\alpha = 0.05$ RK-ttest, WSTD and Tt-test in that order are more robust in a mixture distribution. Meanwhile, at 0.01 level of significance, RK-ttest and WSTD in this order are more robust in a mixture distribution.

In summary, further counts at all alpha levels as indicated in Table 7 and Figure 7 revealed that the Rank transformation test is robust and should be used when data set come from mixture distribution.

Consequently, this study has been able to identify some non-parametric and semi-parametric inferential test statistics that are robust when data comes mixture distribution at different sample sizes and levels of significance. Data access is a restriction for this study, as there are currently no real-life data that follows the combination of Gaussian and Cauchy distributions. Also, for future research, the work can be extended to matched-pairs, two samples and more than two sample problems.

Acknowledgements

The authors appreciate the anonymous reviewers for their in-debt review of the paper and professional contributions.

References

[1] C. A. Ayedun, D. O. Omonijo, O. D. Durodola, M. O. Ajibola, C. O. Oloke, R. Kehinde & O. A. Akinjare, "An evaluation of users' satisfaction level with the quality of the office facilities in some selected private universities in Ogun state, Nigeria", The 33^{rd} International Business Information Management Conference (33^{rd} IBIMA) Granada, Spain between 10^{th} and 11^{th} of April 2019.
[2] J. A. Odukoya, O. Fayomi, D. O. Omonijo, M. C. Anyaegbunam & E. I. Olowookere, "An assessment of the psychological undertones to accidents in manufacturing industries", International Journal of Mechanical and Production Engineering Research and Development (IJMPERD) **9** (2019) 545.
[3] J. A. Odukoya, D. O. Omonijo, S. Misra & R. Ahuja, "Information technology in learning institutions: an advantage or a disadvantage?. In: Abraham A., Panda M., Pradhan S., Garcia-Hernandez L., Ma K. (eds) Innovations in Bio-Inspired Computing and Applications", IBICA 2019. Advances in Intelligent Systems and Computing **1180** (2021). https://doi.org/10.1007/978-3-030-49339-4_34
[4] D. O.Omonijo, M. C. Anyaegbunam, E. Okoye, S. O. Nnatu, B. O. Okunlola, V. A. Adeleke, E. I. Olowookere, O. A. Adenuga & P.Olaoye, "The Influence of genital mutilation on women's sexual activities in Oke-Ona, community, Abeokuta, Nigeria", Journal of Educational and Social Research, **8** (2019) 254. https://doi.org/10.2478/jesr-2019-0044
[5] B. L. S. PrakasaRao, *Nonparametric functional Estimation*, 1st Edition (1983).
[6] P. Denys, "Testing mixed distributions when the mixing distribution is known", Conference paper (2008). https://doi.org/10.1007/978-3-642-01044-6_23
[7] C. JinSeo, P. Jin Seok & P. Sang Woo, "Testing for the conditional geometric mixture Distribution", Journal of Economic Theory and Econometrics **29** (20118) 1.
[8] N. Michael, "On a non-parametric recursive estimator of the mixture distribution", The Indian Journal of Statistics San Antonio Conference: Selected articles **64** (2002) 306.
[9] R. Blair, C. Higgins & J. James, "Comparison of the power of the paired samples t-test to that of Wilcoxon's signed-ranks test under various population shapes", Psychological Bulletin **97** (1985) 119. https://doi.org/10.1037/0033-2909.97.1.119
[10] K. Michelle, Mcdougalli & G. D. Rayner, "Robustness to non-normality of various tests for the one sample location problem, Journal of Applied Mathematics and Decision Sciences **8** (2004) 235.
[11] W. Wolfgang, & E. Alexander, "Robustness and power of the parametric t test and the nonparametric Wilcoxon test under non-independence of observations", Psychological Test and Assessment Modelling **55** (2013).
[12] S. A. Adegoke, T.J. Adejumo & K. Ayinde, "A study on sensitivity and robustness of matched-pairs inferential test statistics to outliers", FUTA Journal of Research in Sciences **13** (2017) 350.
[13] K. K. Yuen, "The two-sample trimmed t for unequal population variances", Biometrika, **61** (1974) 165.
[14] H. J. Keselman, F. Wilcox, R.R. Algina & K. A Fradette, "Comparative study of robust tests for spread: asymmetric trimming strategies", British Journal of Mathematical and Statistical Psychology **61** (2008) 235.
[15] A. John (1710), *Sign test-the free encyclopedia*, <http://www.encyclopedia.com (2016).
[16] W. J. Conover & R. Iman, "Rank transformations as a bridge between parametric and non-parametric statistics", The American Statisticians **35** (1981) 125.
[17] F. Wilcoxon, "Individual comparisons by ranking methods", Biometrics Bulletin **1** (1945) 80.
[18] A. Ayinde, J. O. Kuranga & S. G. Solomon, "Empirical investigation of type 1 error rate of some normality test statistics", International Journal of Computer Applications **148** 24.