Evaluating the Robustness of Some Two-Sample inferential Statistics in the Presence of Mixture Distributions: A Simulation study

TAIWO J. ADEJUMO[1,a,*], SUNDAY D. GBOLAGADE[1], OPEYEMI A. OSHUORU[1], SUNDAY O. KOLEOSO[2], OLUWAKAYODE O. SHADARE[1] AND KAMORU T. OYELEKE[3]

[1]Department of Statistics, Ladoke Akintola University of Technology Ogbomoso, NIGERIA
[2]Department of Statistics Federal University of Technology Akure, NIGERIA
[3]Department of Statistics Olabisi Onabanjo University, Ago-Iwoye, NIGERIA

[a]https://orcid.org/0000-0002-9615-8106
[*]Corresponding Author

Abstract: This study investigates the robustness of two-sample inferential statistics when datasets are derived from mixture distributions, where traditional methods like the t-test may fail due to violated assumptions. Using R software, random variables from Standard Normal, Gamma, and Exponential distributions were generated and analyzed using four inferential tests: Rank Transformation t-test (Rt), Wilcoxon Sum Rank Test (WSD and its Asymptotic version WSA), and Trimmed t-test (Tt-test). Robustness was evaluated based on Type I error rates across varying levels of multicollinearity and sample sizes (n=10, 20, 30, 40, 50, 60, 70, 80 and100). A test was deemed robust if it maintained acceptable error rates (α=0.1, 0.05, and 0.01) and demonstrated consistency across multicollinearity levels and sample sizes. At α=0.1, the WSD and Tt-test exhibited the highest robustness. At α=0.05, the Tt-test was the most robust, while at α=0.01, both the Tt-test and WSD were robust, with the Tt-test slightly outperforming. Overall, the Tt-test and WSD consistently demonstrated robustness across all significance levels, suggesting they are reliable alternatives for two-sample problems involving mixture distributions. These findings underscore the importance of selecting robust statistical methods to ensure accurate inferences in complex data scenarios.

Keywords: Mixture Distribution, Inferential Statistics, Non-parametric, Robustness, Probability Distribution

Received: March 14, 2024. Revised: August 22, 2024. Accepted: December 17, 2024. Published: May 15, 2025.

1 Introduction

A mixture distribution refers to the probability distribution of a random variable constructed from a collection of other random variables. These component variables may be random real numbers or vectors, and while they can share the same distributional form, the resulting mixture may be continuous and characterized by a mixture density function. The individual distributions that form the mixture are known as mixture components, and the probabilities associated with each are called mixture weights. In essence, a mixture distribution represents a combination of two or more probability distributions.

In data analysis, violations of the normality assumption are common. These deviations often result from unequal error variances or the presence of outliers. As a result, robust or non-parametric statistical methods are often required. When data do not follow a normal distribution, it indicates that the underlying random variables may not be identically distributed. Such data are typically modeled using mixture distributions, where component distributions may differ in their parameters, contributing to outliers and non-normal characteristics in the dataset.
The concept of mixture distributions has been extensively studied across various disciplines, including biology, the social sciences, engineering, and the physical sciences. [21] conducted foundational work evaluating the

performance of several paired inferential tests—including the paired *t*-test, Wilcoxon signed-rank test, rank transformation *t*-test, and trimmed *t*-test—in the presence of outliers and multicollinearity. His simulations, based on Gaussian data contaminated with outliers, showed that certain robust tests were more effective in handling such irregularities. Subsequently, [18] made significant theoretical contributions to mixture models, forming the basis for both classical and Bayesian statistical applications. [6] introduced the use of a Dirichlet process prior to model unobserved random effects, allowing for unequal variances across sampling units and enabling a smooth non-parametric estimate approximating a Bayesian estimator.

In the early 2000s, a series of empirical studies emerged. [13] examined the mixture hypothesis in geometric distributions using the likelihood ratio test. His simulations revealed relationships between geometric and exponential mixture hypotheses. [14] used Monte Carlo experiments to compare the statistical power of paired parametric and non-parametric tests, finding that test effectiveness varied with context. [8] illustrated the use of mixture distributions in a biological context, showing how fish measurements varied with age and how the overall distribution represented a mixture of age-specific distributions.

In subsequent years, the focus on statistical robustness intensified. [5] assessed the performance of one-sample parametric, semi-parametric, and non-parametric tests in the presence of outliers. [2] extended this research to matched-pair designs, examining robustness under different correlations and sample sizes. Their results indicated that the *t*-test often failed to maintain appropriate Type I error rates. [10] studied mixture models in which the mixing distribution could be identified using Schwarz's Bayesian criterion and Neyman tests. His work introduced smooth goodness-of-fit tests for sequences of independently identically distributed. random variables. That same year, [3], [15], [16] and [17] published reviews and domain-specific research exploring mixture distribution applications across various fields. [1] contributed by evaluating the performance of inferential statistics derived from Gaussian and Cauchy mixtures, recommending the rank transformation test as a robust option across all significance levels, especially in one-sample settings. Most recently [9] highlighted the limitations of the two-sample Hotelling's T^2 test in multivariate analysis, proposing a robust permutation test based on the minimum regularized covariance determinant estimator for high-dimensional data.

In statistical inference, many widely used test statistics—such as the *t*-test and Hotelling's T^2—are based on the assumptions that data are normally and identically distributed. However, real-world data frequently violate these assumptions due to outliers, unequal variances, and structural complexity. These issues are especially common in data generated from mixture distributions, where the underlying random variables come from multiple distinct distributions (e.g., Gaussian, exponential, or Cauchy). Although mixture distributions are prevalent across diverse fields such as biology, engineering, and the social sciences, most robustness studies have focused exclusively on normally distributed data, neglecting more realistic and complex distributional scenarios. This represents a critical gap in understanding how inferential statistics perform under non-normality caused by data drawn from heterogeneous sources.

Moreover, although some robust and non-parametric methods have been developed, their comparative performance across different mixture types, sample sizes, and significance levels remains insufficiently explored. A comprehensive evaluation of both traditional and modern inferential tests in the context of mixture-distributed data is urgently needed.

Despite the growing body of literature, most existing studies have concentrated primarily on data from normal distributions, often overlooking others, such as exponential or heavy-tailed (such as Cauchy) distributions. This study aims to address that gap by examining the robustness of various two-sample inferential test statistics using data drawn from mixture distributions. Specifically, it investigates mixtures of normal, exponential, and Cauchy distributions. The objective is to identify non-parametric and semi-parametric methods that maintain robustness across different sample sizes and significance levels. Detailed simulation procedures and distributional assumptions are presented in the following sections.

2. Materials and methods

2.1 Distributions used for the Study

In this study, data were generated from four distributions, namely; the normal distribution, gamma distribution and the exponential distribution.

(i) Normal distribution:

The normal distribution is the most widely known and used of all distribution and because it can approximate many natural phenomena so well, it has developed into a standard of reference for many probability problems.

Properties of the Normal distribution

i. It is symmetric about the mean and has bell shaped
ii. Its random variable ranges from $-\infty$ to ∞
iii. It has two parameters, μ and σ.

The normal density function is

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

(ii) Gamma Distribution

Gamma distribution is a two-parameter family of continuous probability distributions. The exponential distribution, Erlang distribution, and chi-squared distribution are special cases of the gamma distribution. There are three different parametrizations in common use:

i. With a shape parameter k and a scale parameter θ.
ii. With a shape parameter α = k and an inverse scale parameter β = 1/θ, called a rate parameter.
iii. With a shape parameter k and a mean parameter μ = kθ = α/β.

We say that a random variable X is distributed gamma if

X ~ Gamma (α, β)

$$f(x,\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{(\alpha-1)} \ell^{-\beta x} \quad (2)$$

$0 < x < \infty$, $\alpha > 0$, $\beta > 0$

where, mean = $\frac{\alpha}{\beta}$ and variance = $\frac{\alpha}{\beta^2}$

(iii) Exponential Distribution

A continuous random variable X is said to have an Exponential (λ) distribution if it has probability density function

$$f_X(x|\lambda) = \begin{cases} \lambda e^{-\lambda x}, & for\ x > 0 \\ 0 & for\ x \leq 0 \end{cases} \quad (3)$$

where $\lambda > 0$ is called the *rate* of the distribution. In the study of continuous-time stochastic processes, the exponential distribution is usually used to model the time until something happens in the process. The mean is $1/\lambda$ and the variance is $1/\lambda^2$

2.2 Review of some inferential Statistic

(i) Trimmed t-test for two independent two-samples

[21] proposed the Trimmed t-test for the independent two-sample case, under unequal population variances. The trimmed mean is an attractive alternative to the mean and the median, because it effectively deals with outliers without discarding most of the information in the data set. Research has shown that the use of trimming (and other modern procedures) results in substantial gains in terms of control of Type I error, power, and narrowing confidence intervals [12]. Also, if data are normally distributed, the mean and the trimmed mean will be the same. [5]

In each sample, the trimmed mean is computed by removing g-observations from each tail of the distribution:

Given the Winsorized mean, the Winsorized sum-of-squared derivation is computed as:

$$\mathrm{SSD_w} = [\mathrm{g}+1]\left[\mathrm{x_{g+1}} - \bar{\mathrm{X}}_\mathrm{w}\right]^2 + \left[\mathrm{x_{g+2}} - \bar{\mathrm{X}}_\mathrm{w}\right]^2 + \cdots + [\mathrm{g}+1]\left[\mathrm{x_{n-g}} - \bar{\mathrm{X}}_\mathrm{w}\right]^2 \quad (4)$$

The trimmed t is obtained by dividing the difference between the trimmed means by the estimated standard error of the difference:

$$t = \frac{\bar{\mathrm{X}}_{\mathrm{t1}} - \bar{\mathrm{X}}_{\mathrm{t2}}}{\sqrt{\frac{S_{w1}^2}{n_1-2g} + \frac{S_{w2}^2}{n_2-2g}}} \quad (5)$$

where; $S_{w1}^2 = \frac{SSD_{w1}}{n_1-2g-1}$, $S_{w2}^2 = \frac{SSD_{w2}}{n_2-2g-1}$

The degrees of freedom are obtained from $\frac{1}{df} = \frac{C^2}{n_1-2g-1} + \frac{(1-C)^2}{n_2-2g-1}$

where $C = \frac{\frac{S_{w1}^2}{(n_1-2g-1)}}{\left[\frac{S_{w1}^2}{(n_1-2g-1)}\right] + \left[\frac{S_{w2}^2}{(n_2-2g-1)}\right]}$

(ii) Wilcoxon rank sum test

Wilcoxon rank sum test is a quick and easy test for two independent samples. It is a good alternative test to the t-test when the data don't meet the assumptions of the test. (It is numerically equivalent to the Mann-Whitney U test). This test can also be performed if only rankings (i.e., ordinal data) are available. It tests the null hypothesis that the two distributions are identical against the alternative that the two distributions differ only with respect to the median. In order words, Wilcoxon rank sum test compares two distributions to assess whether one has systematically larger values than the other. The Wilcoxon test is based on the Wilcoxon rank sum test statistic W, which is the sum of the ranks of one of the samples.

Assumptions for Wilcoxon rank sum test:

(i.) Within each samples the observations are independently and identically distributed.
(ii.) The two samples must be independent of each other.
(iii.) The error terms are mutually independent.
(iv.) The shapes and spreads of the distributions are the same.

The procedures:

(i.) Rank all the data values by assigning rank1 to the smallest data, 2 to the next smallest up to the largest.
(ii.) If one group has fewer values than the other e.g., $n_1<n_2$, add the ranks in the smaller group to get the test statistic W. If $n_1= n_2$, add the ranks in the group containing the smallest ranks.
(iii.) Enter the appropriate table for W, based on sample sizes and determine the probability for W.
(iv.) Based on the p-value, reject H_0 or accept H_0.

The rank sum statistic W becomes approximately normal as the two sample sizes increase. The test Z-statistic by standardizing W is;

$$Z = \frac{W-\mu_w}{\sigma_w} \sim N(0,1) \tag{6}$$

where $\mu_w = \frac{n1(N+1)}{2}$, $\sigma_w = \sqrt{\frac{n1n2(N+1)}{12}}$ and $N = n1 + n2$.

p-value for the Wilcoxon test is based on the sampling distribution of the rank sum statistic W when the null hypothesis (no difference in distributions) is true. P-value can be calculated from special tables, software or a normal approximation (with continuity correction).

(iii) Wilcoxon signed rank test (Asymptotic)

Wilcoxon signed-rank test is named after [19] who in a single paper proposed both the test and rank-sum test for two independent samples. The asymptotic distribution of Wilcoxon signed rank test is:

$$T = \frac{T^+ - E_0(T^+)}{\sqrt{V_0(T^+)}} \sim N(0,1) \tag{7}$$

where $E_0(T^+) = \frac{(n+1)}{4}$ and $V_0(T^+) = \sqrt{\frac{n(n+1)(2n+1)}{24}}$

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.

Source of Data

The following parameters were used to generate data for two samples problems with the aid of R-statistical programming package.

i. Sample size(n) = 10, 20, 30, 40, 50, 60, 70, 80 and100
ii. Replications (RR) = 5000
iii. Hypothesized median (md) = 0
iv. Standard deviation (δ) = 1
v. Correlation (ρ) = 0, 0.3, 0.6, 0.9, 0.95 and 0.99
vi. Presented α-level = 0.1, 0.05 and 0.01

Distributions used for Two Samples Problem

The data were generated from the following distributions

i. Normal distribution with mean (μ) = 0 and standard deviation (δ) = 1
ii. Gamma distribution (n, 0.5)
iii. Exponential distribution (n, 0.5)
where n is the sample size.

The Test Statistics used for Two Samples problem

The test statistics used in the two samples problem are as follows:

i. T-test for Rank transformation (Rt) in two sample by [7]
ii. Wilcoxon sum Rank test (Distribution (WSD) and Asymptotic (WSA)) by [19]
iii. Trimmed t-test (Tt) by [21]

2.3 Procedures for Monte Carlo Experiment

The procedures for data generation and estimation of Type I error rate in two samples mixture distribution are as follows:

i. Choose a sample size(n)

ii. Generate random sample size from the distributions under consideration, X ~ N (n, 0, 1) and Gamma distribution (n, 0.5), Y ~ N (n, 0, 1) and Exponential distribution (n, 0.5).

iii. X and Y are now polluted with correlated observations using equations (8) and (9) as in

$$X = \mu_1 + \sigma_1 Z_1 \quad (8)$$

$$Y = \mu_2 + \rho_{12}\sigma_2 Z_1 + \sqrt{m_{22}} Z_2 \quad (9)$$

where $Z_1 \sim N(0,1)$, $Z_2 \sim N(0,1)$, and

$$m_{22} = \sigma_2^2(1-\rho_{12}^2)$$

In this study, $\rho_{12} = \rho = 0$, 0.3, 0.6, 0.9, 0.95 and 0.99.

iv. Combine the data generated in step (ii).

v. Subject the various test statistics and document their p-values

vi. For each inferential test statistics in step(IV) defined as;

$$H_i = \begin{cases} 1, if\ p-value < \propto \\ 0, otherwise \end{cases} \quad (10)$$

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

vii. From step(ii) to (v) repeat up to 5000 times, RR=5000

viii. For each of the inferential statistics, sum the results obtained in step (vi) as in the equation below;

$$H = \sum_{i=1}^{RR} H_i \quad (11)$$

viii. For each of the inferential statistics, divide the result in step (vii) by the number of replications to estimate the type I error of the test statistics as given as follows:

$$K_\propto = \frac{\sum_{i=1}^{RR} H_i}{RR} = \frac{H}{RR} \quad (12)$$

ix. Choose another sample size (n) to work with and repeat step (ii) to step (ix) until all sample sizes are exhausted.

2.4 Examination of Robustness of the Test Statistics

Robustness of the inferential statistics was 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 [2], used by [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 two sample problem including graphical representation are discussed.

Table 1: Simulation Results at 0.1 Level of Significance

		α = 0.1									
		Rt	WSD	WSA	Tt			Rt	WSD	WSA	Tt
rho=0	10	1	0.2002	1	**0.1158**	rho=0.9	10	1	**0.066**	1	**0.0602**
	20	1	0.382	1	0.2988		20	1	0.2174	1	0.2002
	30	1	0.5308	1	0.4586		30	1	0.4474	1	0.4128
	40	1	0.6644	1	0.6028		40	1	0.6768	1	0.6214
	50	1	0.7634	1	0.7206		50	1	0.831	1	0.7806
	60	1	0.8276	1	0.7984		60	1	0.9256	1	0.889
	80	1	0.8798	1	0.8546		80	1	0.9886	1	0.9764
	100	1	0.965	1	0.9592		100	1	0.9994	1	0.9972
		Rt	WSD	WSA	Tt			Rt	WSD	WSA	Tt
rho=0.3	10	1	0.197	1	**0.12**	rho=0.95	10	1	0.0278	1	0.0352
	20	1	0.4286	1	0.3356		20	1	**0.102**	1	**0.114**
	30	1	0.619	1	0.5346		30	1	0.2596	1	0.2582

	40	1	0.7732	1	0.7114		40	1	0.4514	1	0.4384
	50	1	0.8696	1	0.8264		50	1	0.6392	1	0.6054
	60	1	0.9242	1	0.895		60	1	0.7974	1	0.7536
	80	1	0.954	1	0.9356		80	1	0.952	1	0.9258
	100	1	0.996	1	0.991		100	1	0.9948	1	0.9858
		Rt	WSD	WSA	Tt			Rt	WSD	WSA	Tt
rho=0.6	10	1	0.16	1	**0.1128**	rho=0.99	10	1	0.0024	1	0.0038
	20	1	0.4188	1	0.3334		20	1	0.0064	1	0.012
	30	1	0.6538	1	0.573		30	1	0.012	1	0.0228
	40	1	0.8204	1	0.7624		40	1	0.0274	1	0.042
	50	1	0.9118	1	0.873		50	1	**0.0518**	1	**0.0766**
	60	1	0.963	1	0.9402		60	1	**0.101**	1	**0.1264**
	80	1	0.9936	1	0.9872		80	1	0.2548	1	0.2592
	100	1	0.9996	1	0.998		100	1	0.4678	1	0.4312

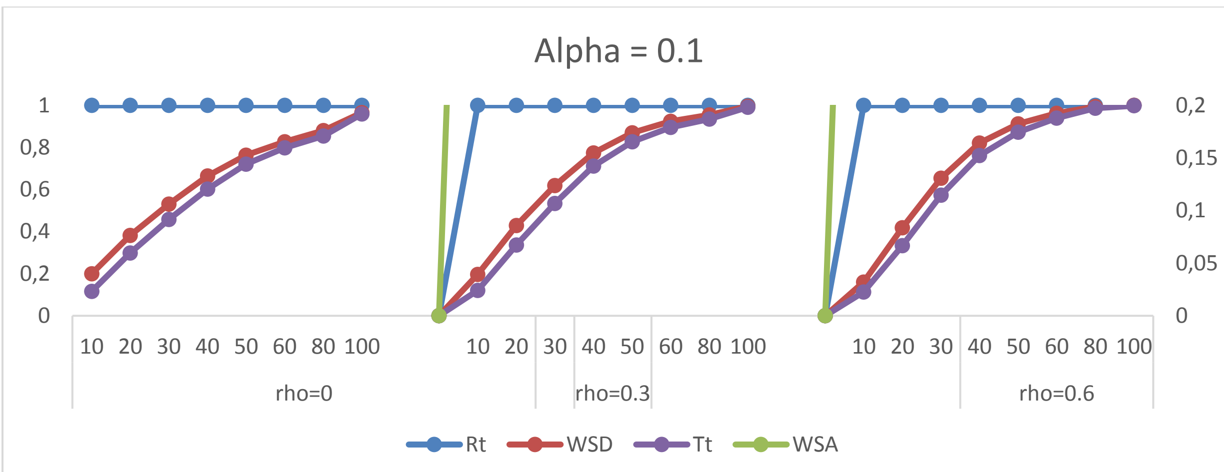

Figure 1a: Graphical Representation of Type I Error rate of Two Sample Statistics in Mixture Distribution across Levels of Multicollinearity and Sample Sizes when α = 0.1

Table 2: Times Type I Error rates approximate to α = 0.1, 0.05 and 0.01

$\alpha = 0.1$

Test Statistics	10	20	30	40	50	60	80	100	SUM	RANK
Rt	0	0	0	0	0	0	0	0	0	3.5
WSD	1	1	0	0	1	1	0	0	4	2
WSA			0	0			0	0	0	3.5
Tt	4	1	0	0	1	1	0	0	7	1

$\alpha = 0.05$

Test Statistics	10	20	30	40	50	60	80	100	SUM	RANK
Rt	0	0	0	0	0	0	0	0		3
WSD	0	0	0	0	0	0	0	0		3
WSA	0	0	0	0	0	0	0	0		3
Tt	3	0	0	0	0	0	0	0	3	1

$\alpha = 0.01$

Test Statistics	10	20	30	40	50	60	80	100	SUM	RANK
Rt	0	0	0	0	0	0	0	0		3.5
WSD	1	2	0	1	0	0	0	0	4	2
WSA	0	0	0	0	0	0	0	0		3.5
Tt	2	2	1	0	0	0	0	1	6	1

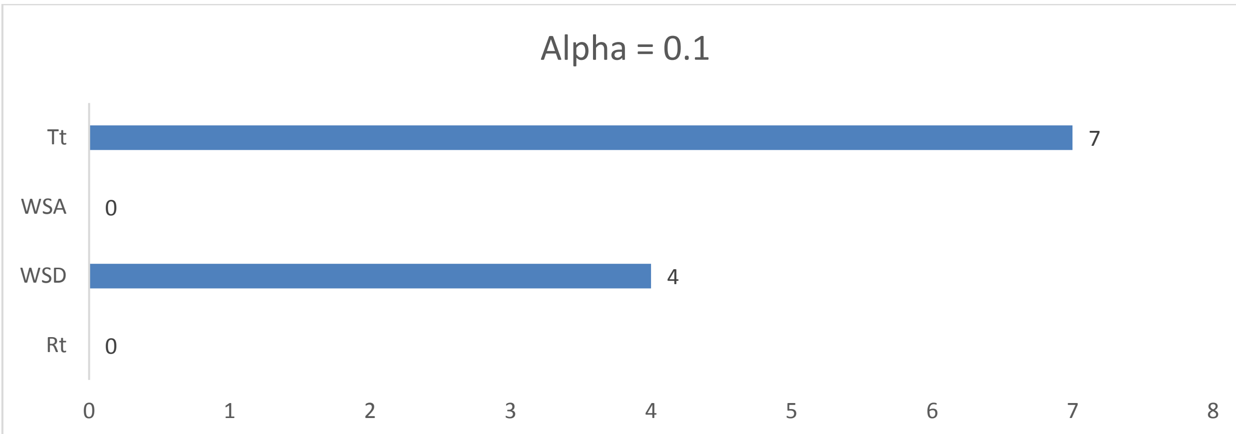

Figure 1b. Bar Chart Indicating Total Times Type I Error rates approximate to α = 0.1

Table 3: Two Sample Simulation Result at 0.05 Level of Significance

$\alpha = 0.05$

		Rt	WSD	WSA	Tt			Rt	WSD	WSA	Tt
rho = 0	10	1	0.109	1	**0.048**	rho = 0.9	10	1	0.0196	1	0.0214
	20	1	0.255	1	0.1678		20	1	0.0776	1	0.0802
	30	1	0.3942	1	0.3018		30	1	0.2098	1	0.2004
	40	1	0.517	1	0.4352		40	1	0.396	1	0.3694
	50	1	0.6278	1	0.5552		50	1	0.5898	1	0.5406
	60	1	0.714	1	0.655		60	1	0.7526	1	0.7074
	80	1	0.7838	1	0.7352		80	1	0.933	1	0.8976
	100	1	0.926	1	0.9086		100	1	0.992	1	0.9798
		Rt	WSD	WSA	Tt			Rt	WSD	WSA	Tt
rho = 0.3	10	1	0.0984	1	**0.0492**	rho = 0.95	10	1	0.0062	1	0.0094
	20	1	0.273	1	0.1828		20	1	0.0278	1	0.0386
	30	1	0.4486	1	0.3518		30	1	0.0816	1	0.102
	40	1	0.6198	1	0.5258		40	1	0.1824	1	0.2036
	50	1	0.7468	1	0.667		50	1	0.3258	1	0.3342
	60	1	0.8406	1	0.7836		60	1	0.4898	1	0.4942
	80	1	0.8952	1	0.8564		80	1	0.7782	1	0.7308
	100	1	0.9832	1	0.9724		100	1	0.9378	1	0.9024
		Rt	WSD	WSA	Tt			Rt	WSD	WSA	Tt
rho = 0.6	10	1	0.0728	1	0.0442	rho = 0.99	10	1	0	1	6.00E-04
	20	1	0.2376	1	0.1658		20	1	6.00E-04	1	0.002
	30	1	0.4488	1	0.3564		30	1	0.0018	1	0.0048
	40	1	0.6504	1	0.5608		40	1	0.0032	1	0.0092

50	1	0.7946	1	0.7174	50	1	0.0056	1	0.0148
60	1	0.89	1	0.8382	60	1	0.0136	1	0.0306
80	1	0.9738	1	0.9528	80	1	0.0438	1	0.0744
100	1	0.9958	1	0.9894	100	1	0.1202	1	0.1506

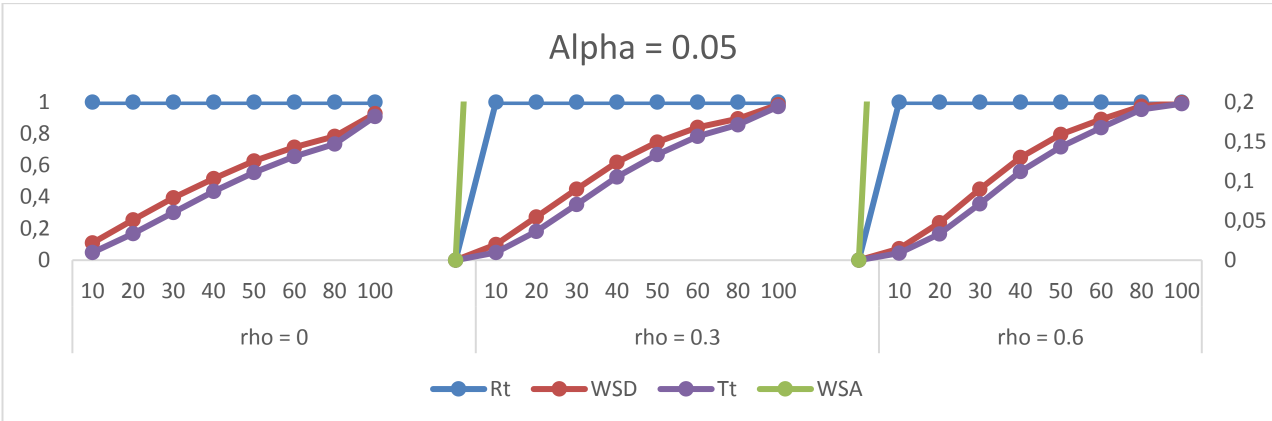

Figure 2a: Graphical Representation of Type I Error rate of Two Sample Statistics in Mixture Distribution across Levels of Multicollinearity and Sample Sizes when α = 0.05

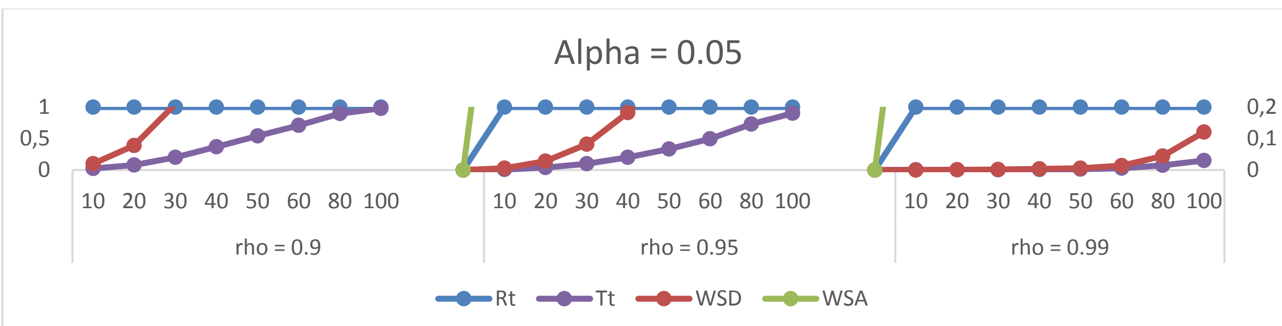

Figure 2b: Graphical Representation of Type I Error rate of Two Sample Statistics in Mixture Distribution across Levels of Multicollinearity and Sample Sizes when α = 0.05

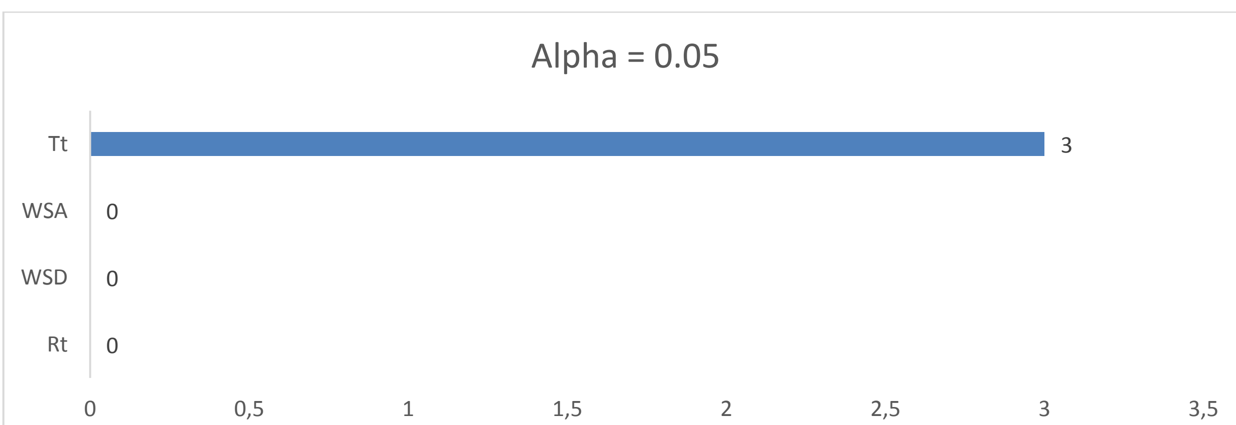

Figure 2c. Bar Chart Indicating Total Times Type I Error rates approximate to α = 0.05

Table 4: Two Sample Simulation Result at 0.01 Level of Significance

Alpha = 0.01											
		Rt	WSD	WSA	Tt			Rt	WSD	WSA	Tt
Rho=0	10	1	0.0214	1	**0.0074**	Rho=0.9	10	1	0.0012	1	0.0022
	20	1	0.073	1	0.0346		20	1	**0.0056**	1	**0.0086**
	30	1	0.1654	1	0.1028		30	1	0.0168	1	0.023
	40	1	0.291	1	0.1904		40	1	0.053	1	0.0688
	50	1	0.4142	1	0.3018		50	1	0.122	1	0.1394
	60	1	0.5406	1	0.4348		60	1	0.223	1	0.2386
	80	1	0.6514	1	0.5474		80	1	0.5082	1	0.479
	100	1	0.8878	1	0.8332		100	1	0.762	1	0.7178
		Rt	WSD	WSA	Tt			Rt	WSD	WSA	Tt
Rho=0.3	10	1	**0.0092**	1	**0.0052**	Rho=0.95	10	1	2.00E-04	1	8.00E-04
	20	1	0.0482	1	0.028		20	1	4.00E-04	1	0.003
	30	1	0.1316	1	0.0878		30	1	0.0032	1	**0.0082**
	40	1	0.2584	1	0.1888		40	1	**0.0104**	1	0.0218
	50	1	0.4026	1	0.3114		50	1	0.0266	1	0.0542
	60	1	0.5596	1	0.4706		60	1	0.0592	1	0.0942
	80	1	0.8108	1	0.7292		80	1	0.197	1	0.2428
	100	1	0.9316	1	0.8846		100	1	0.426	1	0.4488
		Rt	WSD	WSA	Tt			Rt	WSD	WSA	Tt
Rho=0.6	10	1	0.0012	1	0.0022	Rho=0.99	10	1	0	1	0
	20	1	**0.0056**	1	**0.0086**		20	1	0	1	2.00E-04
	30	1	0.0168	1	0.023		30	1	0	1	0
	40	1	0.053	1	0.0688		40	1	0	1	4.00E-04
	50	1	0.122	1	0.1394		50	1	2.00E-04	1	6.00E-04
	60	1	0.223	1	0.2386		60	1	0	1	8.00E-04
	80	1	0.5082	1	0.479		80	1	0	1	0.0028
	100	1	0.762	1	0.7178		100	1	0.001	1	**0.0082**

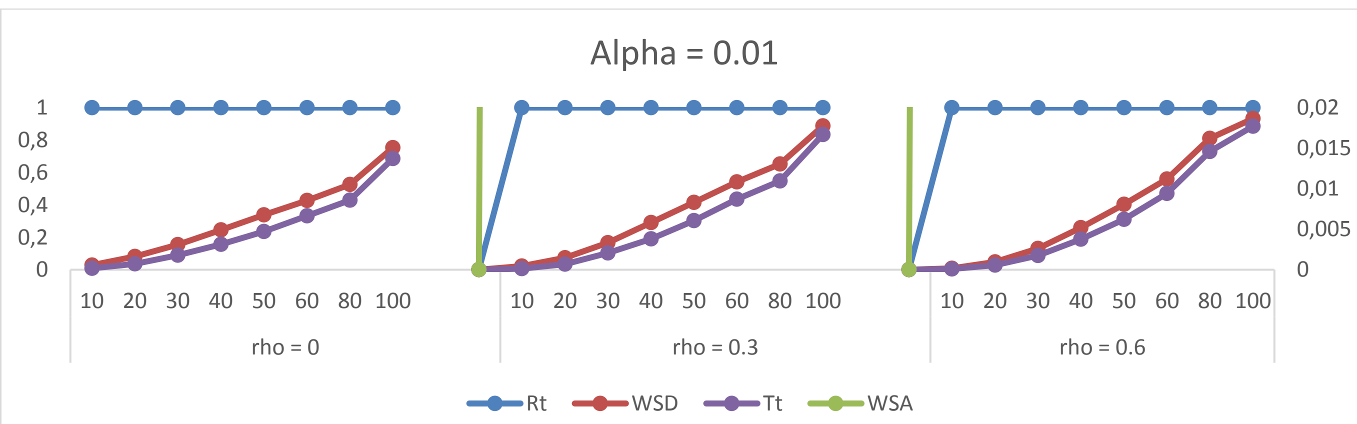

Figure 3a: Graphical Representation of Type I Error rate of Two Sample Statistics in Mixture Distribution across Levels of Multicollinearity and Sample Sizes when α = 0.01

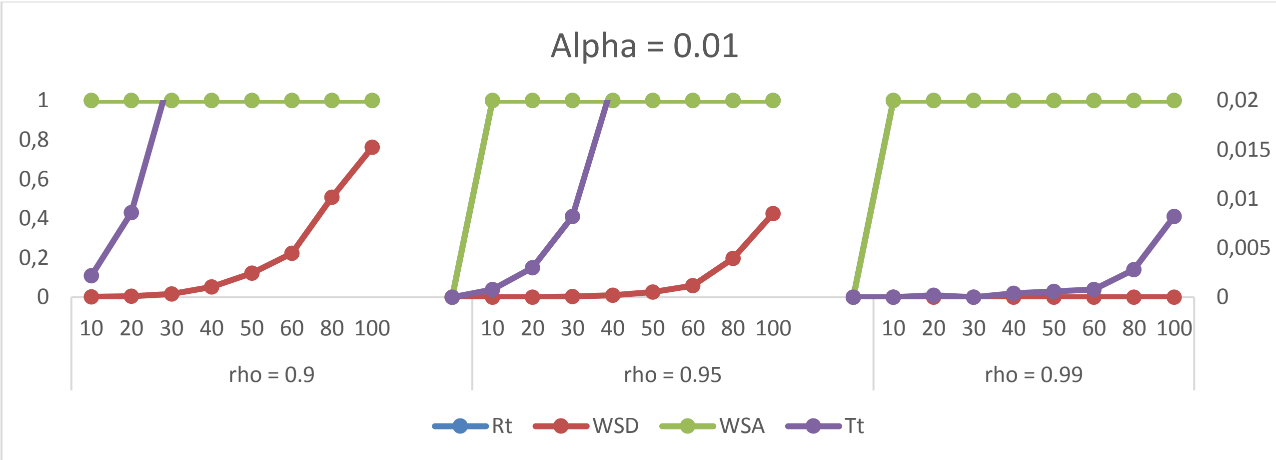

Figure 3b: Graphical Representation of Type I Error rate of Two Sample Statistics in Mixture Distribution across Levels of Multicollinearity and Sample Sizes when α = 0.01

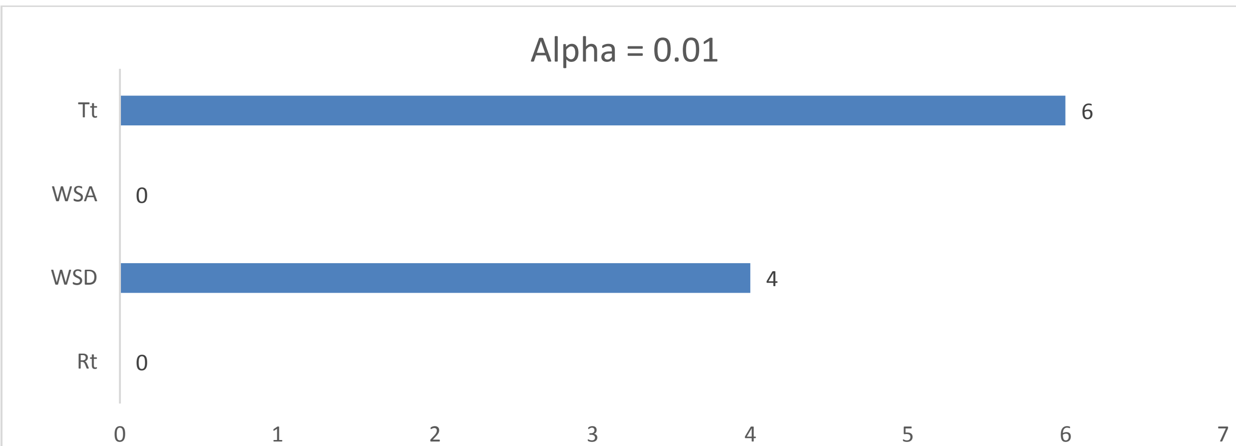

Figure 3c. Bar Chart Indicating Total Times Type I Error rates approximate to α = 0.01

4. Discussion

The simulation results for Type I error rates of two-sample inferential tests, as presented in Table 1 and graphically depicted in Figures 1a and 1b, revealed the following: at α=0.1, the Tt-test and the WSD, in this order, exhibit superior Type I error rates as multicollinearity and sample sizes increase, while the Rt and WSA tests show lower Type I error rates. Furthermore, when aggregated across all levels of multicollinearity for each sample size, as shown in Table 2 and Figure 1c, the Tt-test performs better than the other tests at the α=0.1significance level. Similarly, the results for α=0.05, presented in Table 3 and illustrated in Figures 2a and 2b, indicate that only the Tt-test maintains superior Type I error rates as multicollinearity and sample sizes increase. Aggregated results across all levels of multicollinearity for each sample size, as depicted in Table 2 and Figure 2c, further confirm that the Tt-test outperforms all other test statistics considered in the study. At α=0.01, as shown in Table 4 and graphically in Figures 3a and 3b, the Tt-test and WSD, in this order, achieve better Type I error rates as multicollinearity and sample sizes increase. When aggregated across all multicollinearity levels for each sample size, as shown in Table 4 and Figure 3c, the Tt-test emerges as the top performer at the α=0.01 significance level.

Overall, the investigation of simulation results for two-sample inferential statistics across different significance levels and multicollinearity conditions, as detailed in Tables 1, 3, and 4 and graphically represented in Figures 1, 2, and 3 revealed that correlation really affects the performance of all the test statistics considered as some of them were unable to maintain the pre-selected alpha level excepts Tt test at each sample size. Hence, in the long run, the Tt-test outperformed other test statistics at each sample sizes as it has the highest counts. Additionally, the frequency with which the Type I error rates of the test statistics fall within the preferred interval has been summarized in Table 6. This table ranks the robustness of the two-sample inferential statistics for mixture distributions in order of importance.

Table 5. Total number Times Type I error rate approximates to true error rates when counted across the sample sizes

Test Statistics	10	20	30	40	50	60	80	100	SUM	RANK
Rt	0	0	0	0	0	0	0	0	0	3.5
WSD	2	3	0	1	1	1	0	0	8	2
WSA	0	0	0	0	0	0	0	0	0	3.5
Tt	6	3	1	0	1	1	0	1	13	1

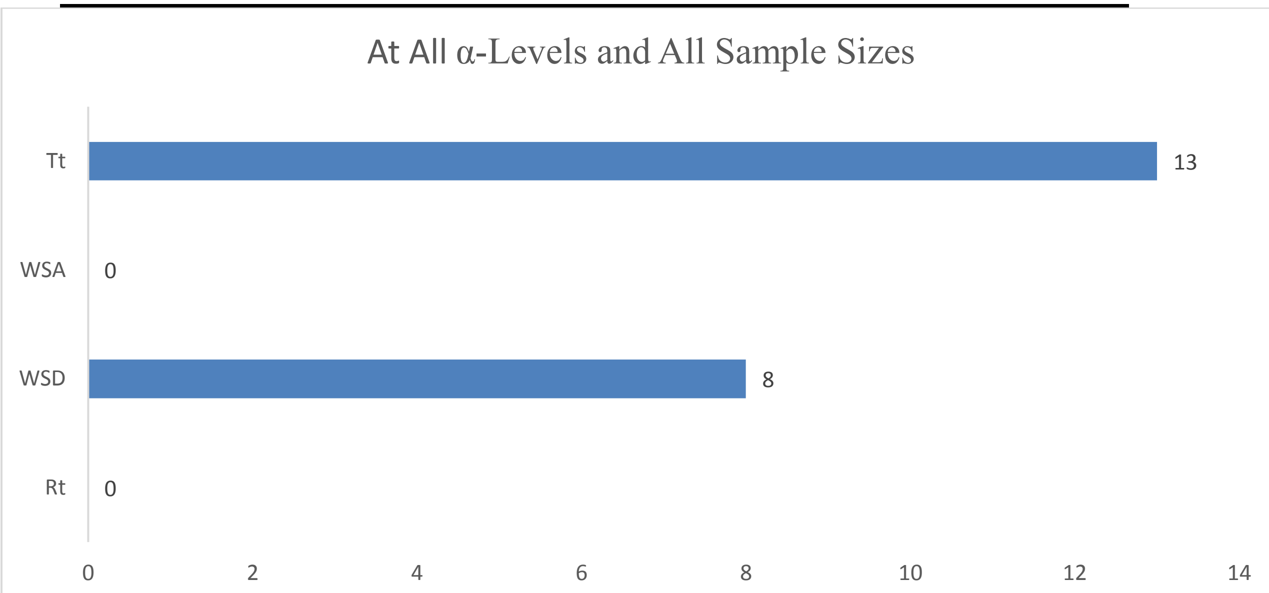

Figure 4. Bar chart indicating overall total times Type I error rates approximates to the true error rates across all sample sizes

Table 6. Overall Summary of Robustness of the inferential Statistics in Mixture Distribution

Alpha Level	Test statistics
0.1	WSD and Tt
0.05	Tt
0.01	Tt and WSD
Overall	Tt and WSD

5. Conclusion

The simulation results demonstrate that the Trimmed t-test (Tt-test) and the Wilcoxon Sum Rank Test (WSD) exhibit robust Type I error rates across varying levels of significance, sample sizes, and multicollinearity in mixture distributions. When results are aggregated across all levels of multicollinearity and sample sizes, the Tt-test consistently demonstrates superior robustness, with the

WSD also performing reliably in certain conditions. These findings, summarized in Table 5, provide a clear ranking of robustness for the test statistics in mixture distributions, highlighting the effectiveness of the Tt-test as the most reliable option across the evaluated conditions. Overall, this study underscores the importance of selecting robust inferential statistics like the Tt-test and WSD for accurate hypothesis testing in complex data scenarios, such as mixture distributions, particularly when standard assumptions are not met.

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

References

[1] Adejumo, T.J., Akomolafe, A.A., Okegbade, A.I. and Gbolagade, S.D. (2022) "A Simulation Study on Robustness of One Sample Inferential Statistics in Mixture Distribution", *African Scientific Reports* 1(142-153)

[2] Ajiboye. A. S, Adejumo, T.J. and Ayinde, K. (2017). A study on sensitivity and robustness of matched-pairs inferential test statistics to outliers, *FUTA Journal of Research in Sciences* 13;350.

[3] Ayedun, C.A., Omonijo, D.O., Durodola, O.D., Ajibola, M.O., Oloke, C.O., Kehinde, R. and Akinjare, O.A. (2019). An evaluation of users' satisfaction level with the quality of the office facilities in some selected private universities in Ogun state, Nigeria, *The 33rd International Business Information Management Conference (33rd IBIMA) Granada*.

[4] Ayinde, K., Kuranga, J.O. and Solomon, S.G. (2009). Empirical investigation of type 1 error rate of some normality test statistics, *International Journal of Computer Applications* 148;24.

[5] Ayinde, K., Adejumo, T. J. and Solomon, G. S. (2016). A Study on Sensitivity and Robustness of One Sample Test Statistics to Outliers. *Global Journal of Science Frontier Research: F Mathematics and Decision Science* Vol. 16 (6).

[6] Blair, R., Higgins, C. and James, J. (1985). 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;119.https://doi.org/10.1037/0033-2909.97.1.119*

[7] Conover, W.J. and Iman, R. (1981). Rank transformations as a bridge between parametric and non-parametric statistics, *The American Statisticians* 35;125.

[8] Denys, P. (2008). Testing mixed distributions when the mixing distribution is known, *Conference paper. https://doi.org/10.1007/978-3-642-01044-623*

[9] Hasan, B. Soofia, I. Nosheen, F. and Olayan, A. (2024). A Robust High-Dimensional Test for Two-Sample Comparisons. *Axioms. 13*(9), 585; https://doi.org/10.3390/axioms13090585.

[10] JinSeo, C. Jin, P. Seok and Sang Woo, P. (2018). Testing for the conditional geometric mixture Distribution, *Journal of Economic Theory and Econometrics* 29;1.

[11] John, A. (2016). Sign test-the *free encyclopedia, https://www.encyclopedia.com*

[12] Keselman, H.J., Wilcox, F., Algina, R.R. and Fradette, K.A. (2008). Comparative study of robust tests for spread: asymmetric trimming strategies, *British Journal of Mathematical and Statistical Psychology* 61;235.

[13] Michael, N. (2002). On a non-parametric recursive estimator of the mixture distribution, *The Indian Journal of Statistics San Antonio Conference*: Selected articles 64;306.

[14] Michelle, K., Mcdougalli and Rayner, G.D. (2004). Robustness to non-normality of various tests for the one sample location problem, Journal of Applied Mathematics and Decision Sciences 8;235.

[15] Odukoya, J.A., Fayomi, O., Omonijo, D.O., Anyaegbunam, M.C. and Olowookere, E.I. (2019). An assessment of the psychological undertones to accidents in manufacturing industries, *International Journal of Mechanical and Production Engineering Research and Development (IJMPERD)* 9; 545.

[16] Odukoya, J.A., Omonijo, D.O, Misra, S. and Ahuja, R. (2019). 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. *Advances in Intelligent*

Systems and Computing 1180 (2021). https://doi.org/10.1007/978-3-030-49339-4 34

[17] Omonijo, D.O., Anyaegbunam, M.C., Okoye, E., Nnatu, S.O., Okunlola, B.O., Adeleke, V.A., Olowookere, E.O, Adenuga, A.O, and Olaoye, P. (2019). The Influence of genital mutilation on women's sexual activities in Oke-Ona, community, Abeokuta, Nigeria, *Journal of Educational and Social Research, 8; 254. https://doi.org/10.2478/jesr-2019-0044*

[18] Prakasa Rao, B.L.S. Nonparametric functional Estimation, 1st Edition (1983).

[19] Wilcoxon, F. (1945). Individual comparisons by ranking methods, *Biometrics Bulletin* 1;80.

[20] Wolfgang, W., and Alexander, E. (2013). Robustness and power of the parametric t test and the nonparametric Wilcoxon test under non-independence of observations, *Psychological Test and Assessment Modelling 55.*

[21] Yuen, K.K. (1974). The two-sample trimmed t for unequal population variances, *Biometrika*, 61;165.

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)

The authors equally contributed in the present research, at all stages from the formulation of the problem to the final findings and solution.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself

No funding was received for conducting this study.

Conflict of Interest

The authors have no conflicts of interest to declare that are relevant to the content of this article.