

Comparison of Type I Error Rates of Siegel-Tukey and Savage Tests among Non-Parametric Tests

Sahib Ramazanov¹, Hakan Çora²

¹Department of Business Administration, Graduate Education Institute, Karabük University, Karabük, Türkiye

²Department of International Relations, Faculty of Business and Management Sciences, Istanbul Okan University, Istanbul, Türkiye
Email: sahibramazanli@gmail.com, corahakan@gmail.com

How to cite this paper: Ramazanov, S. and Çora, H. (2024) Comparison of Type I Error Rates of Siegel-Tukey and Savage Tests among Non-Parametric Tests. *Open Journal of Applied Sciences*, 14, 2393-2410.

<https://doi.org/10.4236/ojapps.2024.149158>

Received: August 19, 2024

Accepted: September 7, 2024

Published: September 10, 2024

Copyright © 2024 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

This study aimed to examine the performance of the Siegel-Tukey and Savage tests on data sets with heterogeneous variances. The analysis, considering Normal, Platykurtic, and Skewed distributions and a standard deviation ratio of 1, was conducted for both small and large sample sizes. For small sample sizes, two main categories were established: equal and different sample sizes. Analyses were performed using Monte Carlo simulations with 20,000 repetitions for each scenario, and the simulations were evaluated using SAS software. For small sample sizes, the I. type error rate of the Siegel-Tukey test generally ranged from 0.045 to 0.055, while the I. type error rate of the Savage test was observed to range from 0.016 to 0.041. Similar trends were observed for Platykurtic and Skewed distributions. In scenarios with different sample sizes, the Savage test generally exhibited lower I. type error rates. For large sample sizes, two main categories were established: equal and different sample sizes. For large sample sizes, the I. type error rate of the Siegel-Tukey test ranged from 0.047 to 0.052, while the I. type error rate of the Savage test ranged from 0.043 to 0.051. In cases of equal sample sizes, both tests generally had lower error rates, with the Savage test providing more consistent results for large sample sizes. In conclusion, it was determined that the Savage test provides lower I. type error rates for small sample sizes and that both tests have similar error rates for large sample sizes. These findings suggest that the Savage test could be a more reliable option when analyzing variance differences.

Keywords

Non-Parametric Test, Siegel-Tukey Test, Savage Test, Monte Carlo Simulation, Type I Error

1. Introduction

As an important component of modern scientific research, statistics has assumed an increasingly critical role, with various statistical analysis methods widely applied, especially in the social and health sciences. Statistical science uses samples to make sense of complex population data and to generate reliable estimates from these data. Among the fields of applied statistics, biostatistics, biometrics, econometrics, sociometrics and archaeometrics offer a wide range of interdisciplinary applications and are used extensively in these fields. Especially in medical research, statistical methods are used intensively and the findings obtained from these methods are highly accepted in the scientific community. This shows how important and indispensable a role statistical science plays in research as a methodological foundation.

The majority of statistical methods have two key features. First, it is assumed that the underlying density function used in the analysis is known. This assumption implies that the distribution of the population to which the sample belongs is known. Second, the focus is on testing hypotheses about the parameters of the density function or their estimates. Such tests are called parametric tests and are based on certain population assumptions. For example, the assumptions that the population is normally distributed and that samples are randomly selected are the mainstays of parametric tests. However, when these assumptions are not met, nonparametric techniques must be used. Nonparametric methods provide more flexible and reliable results when distributional assumptions are not valid. These methods have been proposed since the late 1940s and have been widely researched since then.

Forecasting theory is an effective statistical method that aims to predict future events based on available data. This theory plays an important role in decision-making processes in various disciplines. For example, in fields such as economics, engineering, health sciences and social sciences, forecasting models are used to make informed decisions about future trends and events. Statisticians develop effective forecasting models by assessing the suitability of data sets and the accuracy of the methods used. These models are built by accurately collecting, analyzing and interpreting data. Forecasting theory provides decision makers with valuable information to make informed choices based on current data and offers important insights into future trends.

With continuous advances in statistical science, the importance of forecasting theory is increasing. New data collection techniques, more powerful analysis methods and advanced computer technologies allow for a more effective and widespread use of forecasting theory. Therefore, statistical forecasting methods have become an indispensable part of modern scientific research and applications. Statistical methods play a critical role in analyzing and interpreting data through both parametric and nonparametric approaches. Prediction theory is one of the most important application areas of these methods and supports decision-making processes in various disciplines. Studies in these areas reveal the broad and

dynamic nature of statistical science.

Among the non-parametric tests used to detect scale differences, there are two important methods that stand out in the literature: Savage (SAV) test and Siegel-Tukey (ST) test. These tests stand out with their ability to assess differences between scales without relying on parametric distribution assumptions. Due to these features, they can provide reliable results even when the data sets do not fit a particular distribution.

The Siegel-Tukey test developed by Siegel and Tukey (1960) is a method used to determine the variance differences between groups. This test takes an approach to assessing variance differences by assuming that the means or medians of the groups are approximately equal. That is, the test is based on the assumption that the central tendencies of the groups are close to each other in order to understand whether there is a significant difference in variance between the groups [1]. The test developed by Siegel and Tukey (1960) has a remarkable place among ranking tests and can be applied using mean and median compatible data. Research has shown that the Siegel-Tukey test exhibits a distribution-independent property on samples taken from the same distribution. This shows that the actual alpha level of the test is close to the nominal alpha level and thus provides reliable results [2].

Referring to the simulation-based approach in the study by Tuğran and colleagues (2015) provides a consistent assessment with the methods in the existing literature. The work of Tuğran and associates comprehensively examined the correlation coefficients' Type I error rate and power, offering important insights into how such simulation approaches can be used to evaluate the performance of tests [3].

The study by Mukasa and colleagues (2021) thoroughly addressed the impact of parametric and non-parametric tests on business decision-making processes. This research highlights the significance of test performance in real-world applications by examining the performance of variance tests with small sample sizes. In this context, referencing the findings of Mukasa and associates when evaluating test performance with small sample sizes in your study can reinforce your understanding of the practical effects and applications of the tests [4].

If one wants to nonparametrically test the hypothesis H_0 against the hypotheses H_1 or H_2 in the univariate case ($p = 1$), the Siegel-Tukey test may be recommended. The Siegel-Tukey test is defined as a distribution-independent procedure, which is valid for various data distributions [5].

Combined sample $x_1 \leq x_2 \leq \dots \leq x_n$ a data set sorted as ($n = n_1 + n_2$) the variation series generated from this dataset is rearranged in a special way. The rearrangement is handled sequentially, starting from the beginning and end of the series.

$$x_1, x_2, x_n - 1, x_2, x_3, x_n - 2, x_n - 3, x_4, x_5, \dots$$

First, the smallest value x_1 is taken, then the largest value x_n is selected. Then, taking a step back $x_n - 1$ and going one step further x_2 is taken. This process continues between the elements at the beginning and the end of the array

respectively, *i.e.* it is similar to an “inversion” process that continues by approaching the middle of the array. In this way, the rest of the array is “inverted” as the values at each end are sorted [6].

H_0 : The Siegel-Tukey test used to test the hypothesis $\mu_1 = \mu_2$ adopts an approach in which the rankings are adjusted to the true medians. In this methodology, the ranking of the samples is done in a certain order, and this ordering is applied by numbering the sample from both ends. It is illustrated as follows:

$$1 \ 5 \ 8 \ 9 \ \dots \ 7 \ 6 \ 3 \ 2$$

The Siegel-Tukey test performed the rankings with this arrangement. $T(X, Y)$ Expression, X . The sum of the rankings belonging to the group, and in the Siegel-Tukey test, this sum determines the minimum value that the statistic can reach. This is an important parameter that affects the sensitivity and significance of the test’s analysis of the ranking data. In this context,

$$T(X, Y) (= m(m+1)/2)$$

formula is a standard method for calculating the minimum value of the sum of rankings. This formula serves as a basic calculation tool for organizing and evaluating ranking data. This minimum value represents the lowest statistical sum that can be obtained in a given data set of the Siegel-Tukey test and plays an important role in assessing the power and accuracy of the test [7].

Siegel-Tukey test, $X \cup Y$. The ranking is performed by ordering the observations in the combined dataset from smallest to largest. In this sorting process, the smallest observation is assigned rank 1, the largest observation is assigned rank 2, the next largest observation is assigned rank 3 and the next smallest observation is assigned rank 4, and this process is applied to the entire sort. The Siegel-Tukey statistic is calculated based on this ranking order. The null hypothesis is rejected when the value of the statistic is too small or too large. This method offers a rank-based approach to assessing differences between observations and considers the fit of the rank statistics to a given distribution when testing the validity of the null hypothesis [8].

Simple formulas for calculating the Siegel-Tukey test are presented. R_j refers to the specific rank sum used in the Siegel-Tukey procedure for the group. The Siegel-Tukey test statistic is calculated as follows:

$$S - T = \frac{12}{N(N+1)} \sum_{j=1}^k \frac{R_j^2}{n_j} - 3(N+1)$$

This formula allows the calculation of the test statistic of the Siegel-Tukey test based on rank sums and total sample size. This calculation plays a critical role in determining the impact of the ranking data and the accuracy of the test in the process of applying the test.

The Siegel-Tukey test generally assumes that the locations of the distributions are equal. In contrast, the Savage test does not assume that the locations are equal; instead, it recognizes that differences in scale can lead to differences in location. This test works on the assumption that samples are drawn from continuous

distributions. The null hypothesis predicts that there is no difference in spread between the two distributions, while the two-tailed alternative hypothesis suggests that there is a difference in spread.

The test is named after I. R. Savage, Savage, H_0 developed a close ranking test for Lehmann's alternatives to the null hypothesis. However, there are no tables available for practical applications of the Savage test (Šidák, 1973). Savage (1956) developed the now well-known Savage test [9].

Sample 1, x_1, x_2, \dots, x_{n_1} and sample 2, y_1, y_2, \dots, y_{n_2} is defined as. When sorting the merged samples, the sample to which each sample belongs is recorded. x_i sort degree of the value R_i as the test statistic. The test statistic is calculated for both samples.

The test statistic is as follows:

$$S = \sum_{i=1}^m a(R_i)$$

here

$$a(i) = \sum_{j=N+1-i}^N \frac{1}{j}$$

When such sample sizes are sufficiently large, the following normalized approach can be used.

$$S^* = \frac{S - n}{\sqrt{\frac{nm}{N-1} \left(1 - \frac{1}{N} \sum_{j=1}^N \frac{1}{j} \right)}}$$

S^* value is the critical value obtained from the standard normal distribution. z value [10].

2. Material and Method

Simulation Study

Monte Carlo simulation is a powerful method widely used in statistical analysis. The technique is based on repeatedly generating sample data from a known population model. The analyses performed on each sample reveal various properties of the model, such as prediction bias or statistical power. The results obtained in this process are evaluated by averaging over all samples. This provides an overall picture of the model's performance and reliability. Using Monte Carlo simulation in this way facilitates the understanding of complex statistical problems and allows us to better understand how the model behaves under various scenarios [11].

Simulation models based on Monte Carlo methods are based on a large number of repetitions of random experiments [12]. The initial use of these techniques was in the solution of mathematical and statistical problems. For example, numerical determination of the number π , integral calculations of complex functions and estimation of the probability of random events were among the first applications of Monte Carlo methods. This method has played a critical role in improving the

accuracy and reliability of numerical analysis, especially as a powerful tool for problems that are difficult to solve analytically [13].

Monte Carlo simulations were performed using SAS 9.00 software to implement Fleishman's (1978) power function. In this process, Fleishman's power transformation method was used to generate random numbers from a standard normal distribution with zero mean and one standard deviation. SAS's RANNOR procedure was used to generate population distributions. These generated data were transformed through Fleishman's power function equation [14].

$$Y = a + [(d \times X + c) \times X + b] \times X$$

The equation based on Fleishman's power function includes the distribution variable Y and certain constants and is implemented with the normally distributed random variable X with zero mean and one standard deviation. The samples obtained using the RANNOR procedure in SAS software contain constants covering the values a , b , c and d . In this context, a is a constant value and is defined according to the formula $a = -c$, while b , c and d are considered as variable values. Power simulations are performed using the PROC NPAR1WAY procedure once the sample populations have been generated [15].

In this study, we examine the performance of Siegel-Tukey and Savage tests for heterogeneous variances in small and large sample approaches. In planning the test statistics, three different main population distributions, one standard deviation and twenty-four different sample size combinations were considered. Twelve of these combinations were chosen for large sample sizes and the other twelve for small sample sizes.

Sample sizes, n_1 and n_2 expressed as ordered pairs, where n_1 and n_2 represent the sizes of the first and second samples, respectively. Combinations with small and equal sample sizes include 5, 8, 10, 12, 16 and 20, while combinations with small and different sample sizes are (4, 16), (8, 16), (10, 20), (16, 4), (16, 8) and (20, 10). Equal combinations for large sample sizes were 25, 50, 75 and 100, and for large and different sample sizes, combinations of (10, 30), (30, 10), (50, 75), (50, 100), (75, 50), (75, 100), (100, 50) and (100, 75) were used.

The analyses were conducted in SAS software, running 20,000 Monte Carlo simulations for each case. The distributions analyzed include Normal, Platykurtic and Skewed distributions. This comprehensive approach aims to evaluate the effectiveness of Siegel-Tukey and Savage tests under heterogeneous variances for various sample sizes and distribution types.

In comparing the Siegel-Tukey and Savage tests, defining the population distributions is a crucial step. One aspect of the research involves evaluating Type I error rates under different skewness and kurtosis levels deviating from a normal distribution. In this context, identifying populations based on Fleishman's 1978 power function is of significant importance. **Table 1** presents skewness and kurtosis alignments along with the a , b , c , and d coefficients, which have a mean of 0 and a standard deviation of 1, based on Fleishman's 1978 work.

Table 1. Fleishman's power function for $\mu = 0$ and $\sigma = 1$.

Distribution	Skewness (γ_1)	Kurtosis (γ_2)	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>
Normal	0.00	0.00	0.00	1.0000000	0.00	0.00
Platykurtic	0.00	-0.50	0.00	1.0767327	0.00	-0.0262683
Skewed	0.75	0.00	-0.1736300	1.1125146	0.1736300	-0.0503344

a: [16] and [17].

In the study, heterogeneous variances are defined as the situation where the variances of different sample groups in data sets are not equal. The application of heterogeneous variances in simulations focuses on how various distributions (Normal, Platykurtic, Skewed) and the standard deviation ratio represent variance differences. In a normal distribution, heterogeneity is tested by adding variance differences; in a platykurtic distribution, variance differences are accentuated with flattened peaks; while in a skewed distribution, the lack of symmetry more clearly reflects variance heterogeneity. These differences play a critical role in evaluating the performance of Siegel-Tukey and Savage tests. Simulations allow for the analysis of the effects of heterogeneous variances and the sensitivity of tests to these variances, making it possible to understand the effectiveness of tests in variance analysis.

3. Results and Discussion

3.1. Findings Obtained in Small Sample

This research aims to examine the performance of Siegel-Tukey and Savage tests using data sets with heterogeneous variances. The study was conducted by considering three different distribution types and one standard deviation. These distributions are Normal, Platykurtic and Skewed. One standard deviation ratio contains a value such as 1, which represents the differences between the variances.

The analyses focused on small sample sizes and in this context 12 different combinations of sample sizes were used. These sample sizes fall into two main categories:

1) Completely Equal Sample Sizes: These combinations represent cases where all sample sizes are equal. This group includes sample sizes of 5, 8, 10, 12, 16 and 20. Such combinations aim to evaluate the performance of tests in terms of the ability to detect differences in variance between samples of equal size.

2) Different Sample Sizes: These combinations refer to cases where the size differences between two samples are different. This group includes pairs (4, 16), (8, 16), (10, 20), (16, 4), (16, 8) and (20, 10). This variation aims to analyze the ability of tests to assess differences in variance between pairs of samples of different sizes.

These combinations were generated using Monte Carlo simulations. Monte Carlo simulation is a method involving many random replications of a particular experiment or analysis and is used to understand the statistical properties of the

results. For each sample case, 20,000 simulations were performed and processed in SAS software. These large-scale simulations provide a comprehensive assessment of how tests perform under different conditions.

Table 2 shows the type I error rates of the Siegel-Tukey (ST) and Savage (SAV) tests when sample sizes are equal between two samples with small sample sizes. The standard deviation is taken as 1 and the analyses are performed for Normal, Platykurtic and Skewed distributions. Type I error rate refers to the probability that the test falsely rejects the null hypothesis and is usually evaluated at $\alpha = 0.05$.

Table 2. Type I errors of ST and SAV Tests for two samples with equal sample sizes at small sample sizes: standard deviation = 1.

POPULATION DISTRIBUTION	n_1	n_2	TYPE I ERROR RATE	
			ST	SAV
NORMAL	5	5	0.055	0.016*
	8	8	0.051	0.028*
	10	10	0.054	0.036*
	12	12	0.051	0.040*
	16	16	0.045*	0.037*
	20	20	0.047*	0.041*
PLATYKURTIC	5	5	0.054	0.014*
	8	8	0.051	0.028*
	10	10	0.053	0.035*
	12	12	0.051	0.038*
	16	16	0.044*	0.036*
	20	20	0.051	0.045*
SKEWED	5	5	0.056	0.016*
	8	8	0.049*	0.028*
	10	10	0.054	0.036*
	12	12	0.053	0.041*
	16	16	0.046*	0.038*
	20	20	0.047*	0.040*

*Values less than or equal to $\alpha = 0.05$.

Small sample sizes were applied to three different population distributions, with a total of six samples used. Care was taken to ensure that the standard deviation ratios were maintained at 1, and each ratio was calculated individually. The results were presented in graphs, which were detailed according to sample sizes and standard deviation values. The relevant graphs are shown in **Figure 1**, **Figure 2**.

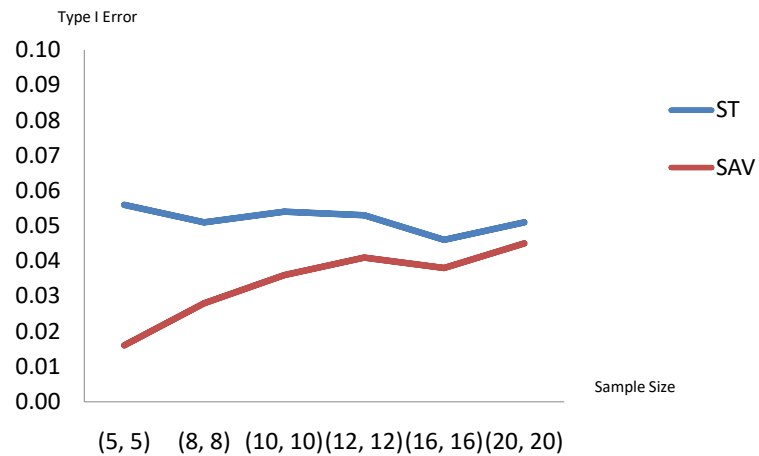


Figure 1. Comparison of Type I error rates of Siegel-Tukey and Savage tests for Normal, Platykurtic and Skewed Distributions with standard deviation set to 1 for small samples, equal sample sizes and heterogeneous variances.

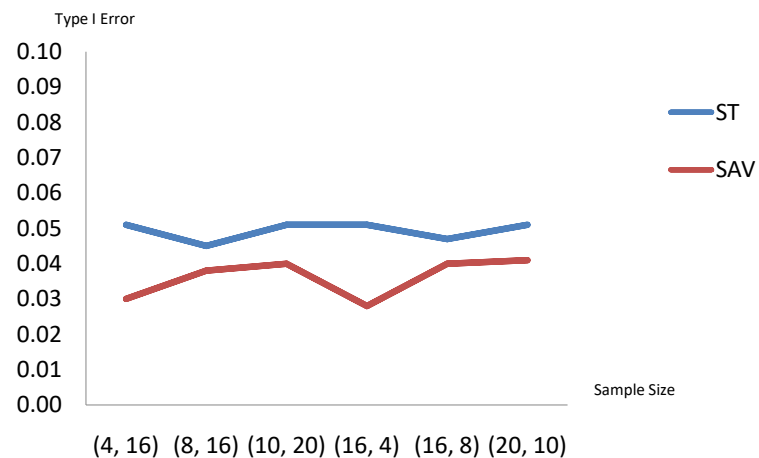


Figure 2. Comparison of Type I error rates of Siegel-Tukey and Savage tests for Normal, Platykurtic and Skewed Distributions when the standard deviation is taken as 1 for small sample sizes, when sample sizes are different and variances are heterogeneous.

Normal Distribution:

- For small sample sizes ($n_1 = n_2$), the type I error rate of the Siegel-Tukey (ST) test generally ranges between 0.045 and 0.055, while the type I error rate of the Savage (SAV) test ranges between 0.016 and 0.041.
- The Siegel-Tukey test slightly decreases the type I error rate as the sample size increases. The lowest rate is observed between 0.045 and 0.047 and is lowest for large sample sizes.
- The Savage test, on the other hand, has lower type I error rates in small sample sizes and up to 0.041 in large sample sizes.

Platykurtic Distribution:

- The type I error rate of the Siegel-Tukey test varies between 0.044 and 0.054. These rates are similarly distributed in small sample sizes, with the lowest rate

observed as 0.044.

- The type I error rate of the Savage test varies between 0.014 and 0.045. Generally lower rates were observed for small sample sizes. For example, the lowest rate is 0.014 and goes up to 0.045 for large sample sizes.

Skewed Distribution:

- The type I error rate of the Siegel-Tukey test varies between 0.046 and 0.056, while that of the Savage test varies between 0.016 and 0.041.
- Type I error rates of the Siegel-Tukey test for skewed distribution vary between 0.046 and 0.056 depending on the sample size and are slightly lower for larger sample sizes.
- The type I error rate of the Savage test, on the other hand, shows lower rates in small sample sizes and increases up to 0.040 in large sample sizes.

Table 3 shows the type I error rates of the Siegel-Tukey (ST) and Savage (SAV) tests for small sample sizes when there are size differences between two samples. The standard deviation was set to 1 and the analyses were performed for Normal, Platykurtic and Skewed distributions. Type I error rate refers to the probability that the test falsely rejects the null hypothesis and is usually evaluated at $\alpha = 0.05$.

Table 3. Type I errors of ST and SAV Tests in the case of different sample sizes between two samples with small samples: standard deviation rate = 1.

POPULATION DISTRIBUTION	n_1	n_2	TYPE I ERROR RATE	
			ST	SAV
NORMAL	4	16	0.049*	0.028*
	8	16	0.043*	0.036*
	10	20	0.048*	0.038*
	16	4	0.051	0.028*
	16	8	0.043*	0.037*
	20	10	0.049*	0.040*
PLATYKURTIC	4	16	0.046*	0.028*
	8	16	0.045*	0.038*
	10	20	0.051	0.040*
	16	4	0.048*	0.027*
	16	8	0.047*	0.040*
	20	10	0.051	0.041*
SKEWED	4	16	0.051	0.030*
	8	16	0.044*	0.037*
	10	20	0.051	0.040*
	16	4	0.047*	0.028*
	16	8	0.043*	0.037*
	20	10	0.049*	0.039*

*Values less than or equal to $\alpha = 0.05$.

Normal Distribution:

- Type I error rates of the Siegel-Tukey test vary between 0.043 and 0.051 depending on the differences in sample sizes. In the analyses with small sample sizes ($n_1 < n_2$), error rates between 0.043 and 0.049 were observed in combinations where the difference in sample size was larger. For example, the lowest rates were 0.043 and 0.049 in the combinations (4, 16) and (8, 16).
- Type I error rates of the Savage test range between 0.027 and 0.041. For small sample sizes, the Savage test generally shows lower type I error rates. In particular, the lowest rates of 0.027 and 0.028 were observed for the combinations (4, 16) and (16, 4).

Platykurtic Distribution:

- Type I error rates of the Siegel-Tukey test range between 0.045 and 0.051. In analyses with small sample sizes, error rates generally range between 0.045 and 0.051. The lowest rates were 0.046 and 0.048, especially for the combinations (4, 16) and (16, 4).
- Type I error rates of the Savage test vary between 0.027 and 0.041. For small sample sizes, the Savage test also exhibits low error rates when the sample sizes are different. In particular, the lowest rates are 0.027 and 0.028 for the combinations (4, 16) and (16, 4).

Skewed Distribution:

- Type I error rates of the Siegel-Tukey test ranged between 0.043 and 0.051. In the skewed distribution, error rates between 0.043 and 0.051 were observed in combinations where the difference between sample sizes was larger. For example, the lowest rates were 0.043 and 0.044 in the combinations (4, 16) and (8, 16).
- Type I error rates of the Savage test vary between 0.027 and 0.041. Type I error rates of the Savage test are generally lower and show more consistent results when analyzed with small sample sizes. Especially for the combinations (4, 16) and (16, 4), the lowest rates are 0.027 and 0.028.

3.2. Findings Obtained in Large Sample

In total, 12 different sample sizes were used, focusing on large sample sizes. Sample sizes were generated by Monte Carlo simulation and each sample size was applied to three different population distributions. In this analysis, the results are evaluated by considering the standard deviation rates in each population.

Large sample sizes are divided into two main categories:

1) Large and Equal Sample Sizes: This group represents situations where sample sizes are perfectly equal. These equal sample sizes are: 25, 50, 75 and 100. These combinations assess situations where both samples are the same size and aim to examine the performance of the tests under these conditions.

2) Large and Different Sample Sizes: This group refers to situations where there are differences in size between two samples. Large and different sample sizes include the following combinations: (10, 30), (30, 10), (50, 75), (50, 100), (75, 50),

(75, 100), (100, 50) and (100, 75). These combinations are intended to assess how different sized sample pairs affect variance differences.

Monte Carlo simulations were used to analyze the type I errors of the Siegel-Tukey and Savage tests under these sample sizes and distribution types. Based on the simulation results, the effectiveness of the tests was evaluated by considering their performance at each sample size and standard deviation values. The analysis examines a large data set, including Normal, Platykurtic and Skewed distributions.

Table 4 shows the type I error rates of Siegel-Tukey (ST) and Savage (SAV) tests when sample sizes are equal between two samples with large sample sizes. The standard deviation was set to 1 and the analyses were performed for Normal, Platykurtic and Skewed distributions. Type I error rate refers to the probability that the tests falsely reject the null hypothesis and is usually evaluated at $\alpha = 0.05$.

Table 4. Type I errors of ST and SAV Tests for equal sample sizes in large samples: standard deviation rate = 1.

POPULATION DISTRIBUTION	n_1	n_2	TYPE I ERROR RATE	
			ST	SAV
NORMAL	25	25	0.049*	0.043*
	50	50	0.048*	0.046*
	75	75	0.049*	0.047*
	100	100	0.047*	0.046*
PLATYKURTIC	25	25	0.048*	0.042*
	50	50	0.049*	0.047*
	75	75	0.049*	0.048*
	100	100	0.052	0.051
SKEWED	25	25	0.051	0.044*
	50	50	0.052	0.051
	75	75	0.049*	0.047*
	100	100	0.046*	0.045*

*Values less than or equal to $\alpha = 0.05$.

Large sample sizes were applied to three different population distributions, involving two distinct sample groups. The first group consisted of four samples with large and equal sample sizes, while the second group included eight samples with large but varied sample sizes. In determining these sample sizes, care was taken to maintain standard deviation ratios at 1, and these ratios were calculated meticulously. The results were presented in detail according to sample sizes and standard deviation values, and this presentation was achieved through graphs. The resulting graphs, which are shown in **Figure 3**, **Figure 4**, facilitate a better understanding of the data. These graphs provide a detailed examination of the effects of sample size values.

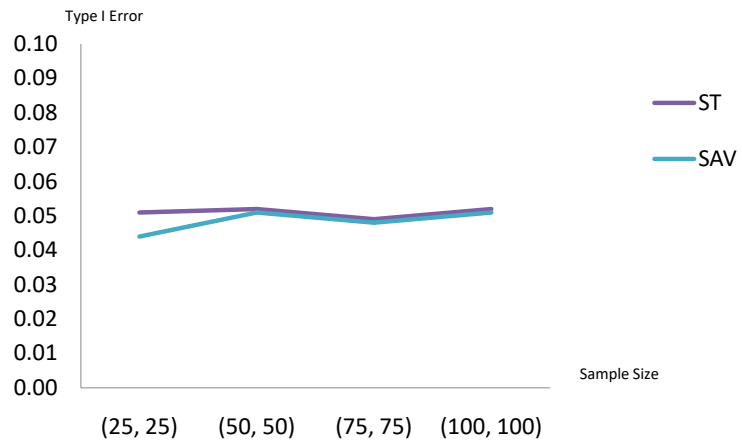


Figure 3. Comparison of Type I error rates of Siegel-Tukey and Savage tests for Normal, Platykurtic and Skewed Distributions when standard deviation is set to 1 for large samples, equal sample sizes and heterogeneous variance conditions.

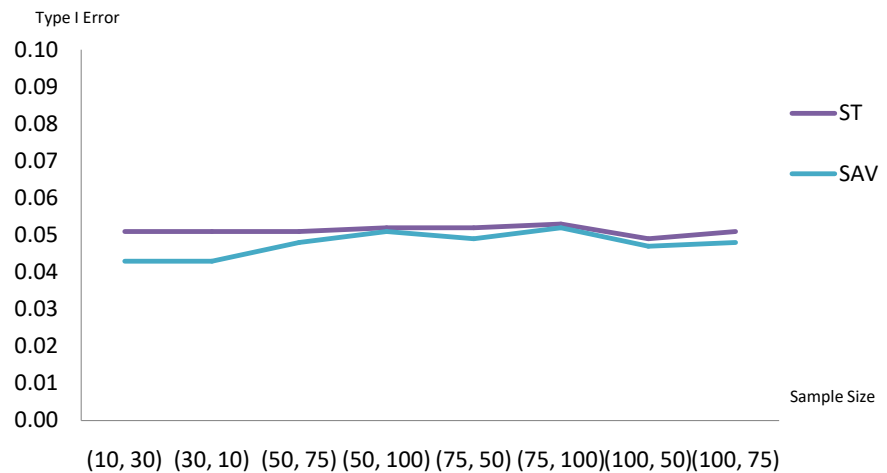


Figure 4. Comparison of Type I error rates of Siegel-Tukey and Savage tests for Normal, Platykurtic and Skewed Distributions with standard deviation set to 1 for large samples, different sample sizes and heterogeneous variance conditions.

Normal Distribution:

- Type I error rates of the Siegel-Tukey test vary between 0.047 and 0.049 as the sample sizes increase. At large sample sizes, the lowest rate was determined as 0.047.
- Type I error rates of the Savage test range from 0.043 to 0.046. For large sample sizes, generally lower values were observed, with the lowest rate being 0.043.

Platykurtic Distribution:

- Type I error rates of the Siegel-Tukey test vary between 0.048 and 0.052 with sample sizes. The highest rate of 0.052 was observed for the sample size of 100.
- Type I error rates of the Savage test range between 0.042 and 0.051. For large sample sizes, the type I error rate is generally higher, with the lowest rate observed at 0.042.

Skewed Distribution:

- Type I error rates of the Siegel-Tukey test vary between 0.046 and 0.052 across sample sizes. For large sample sizes, the lowest rate was 0.046.
- Type I error rates of the Savage test vary between 0.044 and 0.051. The error rates of the Savage test are generally lower for larger sample sizes, with the lowest rate being 0.044.

Table 5 shows the type I error rates of Siegel-Tukey (ST) and Savage (SAV) tests when sample sizes differ between two samples with large sample sizes. The standard deviation was set to 1 and the analyses were performed for Normal, Platykurtic and Skewed distributions. Type I error rate refers to the probability that the tests falsely reject the null hypothesis and is usually evaluated at $\alpha = 0.05$.

Table 5. Type I errors of ST and SAV Tests for different sample sizes in large samples: standard deviation rate = 1.

POPULATION DISTRIBUTION	n_1	n_2	TYPE I ERROR RATE	
			ST	SAV
NORMAL	10	30	0.049*	0.042*
	30	10	0.051	0.042*
	50	75	0.049*	0.047*
	50	100	0.052	0.051
	75	50	0.049*	0.047*
	75	100	0.051	0.049*
	100	50	0.048*	0.047*
	100	75	0.051	0.048*
PLATYKURTIC	10	30	0.048*	0.042*
	30	10	0.051	0.043*
	50	75	0.051	0.048*
	50	100	0.051	0.049*
	75	50	0.052	0.049*
	75	100	0.049*	0.047*
	100	50	0.049*	0.047*
	100	75	0.049*	0.047*
SKEWED	10	30	0.051	0.043*
	30	10	0.048*	0.041*
	50	75	0.051	0.048*
	50	100	0.052	0.051
	75	50	0.048*	0.047*
	75	100	0.053	0.052
	100	50	0.049*	0.047*
	100	75	0.049*	0.048*

*Values less than or equal to $\alpha = 0.05$.

Normal Distribution:

- Type I error rates of the Siegel-Tukey test vary between 0.048 and 0.052 depending on the differences between sample sizes. In analyses with large and different sample sizes, error rates were generally between 0.049 and 0.052. For example, in the (10, 30) and (50, 75) combinations, the rates were 0.049 and 0.049, respectively.
- Type I error rates of the Savage test ranged between 0.041 and 0.051. Type I error rates of the Savage test are generally lower, for example, the lowest rates of 0.041 and 0.042 were observed in the combinations (30, 10) and (10, 30).

Platykurtic Distribution:

- Type I error rates of the Siegel-Tukey test vary between 0.048 and 0.052 when differences in sample sizes are taken into account. These rates were determined as 0.052 and 0.049 for the (75, 50) and (100, 50) combinations, respectively.
- Type I error rates of the Savage test vary between 0.047 and 0.049. When the sample sizes are different, the lowest values of the Savage test error rates are 0.042 and 0.043.

Skewed Distribution:

- Type I error rates of the Siegel-Tukey test ranged between 0.048 and 0.053 for differences in sample sizes. The lowest rates were 0.048 and 0.049, especially for the (10, 30) and (100, 50) combinations.
- Type I error rates of the Savage test ranged between 0.041 and 0.052. The error rates of the Savage test were generally lower, with the lowest rates of 0.041 and 0.043 in the combinations (30, 10) and (10, 30).

The practical implications of preferring the Savage test with small sample sizes highlight the advantages the test provides in certain situations. In small sample sizes, the Savage test's lower Type I error rates indicate that the test offers more reliable results. This enables more accurate results in studies with limited data. The Savage test, which is robust to distributional assumptions, is useful in cases where the normality assumption is not made or there is uncertainty about the distribution. Additionally, reliability is of high importance in small sample sizes; this test ensures accuracy by reducing the risk of erroneously rejecting the null hypothesis. In fields requiring high security and accuracy, such as healthcare and finance, preferring the Savage test can contribute to obtaining more reliable results. Its ability to detect variance differences between samples enhances the test's sensitivity. Overall, the low Type I error rates provided by the Savage test in small sample sizes offer significant advantages in practical applications and make it a tool to be preferred depending on the characteristics of the data set.

4. Conclusions

The findings of this study provide an in-depth examination of the performance of the Siegel-Tukey and Savage tests in data sets with heterogeneous variances. Extensive analysis of the type I error rates of both tests in small and large sample sizes clearly demonstrates their ability to detect variance differences.

The distributions used in the study—Normal, Platykurtic, and Skewed—were chosen to represent different data scenarios. The Normal distribution is the most commonly used distribution in statistical analyses and serves as a reference point for evaluating the fundamental performance of tests. The Platykurtic distribution represents situations with fewer outliers and a wider spread of data, making it important for testing variance differences. The Skewed distribution addresses situations where data symmetry is disrupted and was selected to evaluate asymmetric conditions commonly encountered in real-world data. The chosen 1 standard deviation ratio was determined to make the differences between variances more pronounced, thereby increasing the sensitivity of the tests. These distributions and the standard deviation ratio were used to ensure the validity of analyses across a range of data scenarios and to evaluate the performance of tests in various conditions.

In the analyses performed with small sample sizes, significant differences were observed between the performance of Siegel-Tukey and Savage tests. Under normal distribution conditions, the type I error rates of the Siegel-Tukey test ranged between 0.045 and 0.055, while the type I error rates of the Savage test were between 0.016 and 0.041. This result suggests that the Savage test has lower type I error rates in small sample sizes and is better able to control the probability of falsely rejecting the null hypothesis. Similar trends were observed for the Platykurtic and Skewed distributions, with the Savage test generally providing lower error rates.

In scenarios with different sample sizes, type I error rates of the Savage test were found to be lower than those of the Siegel-Tukey test. These findings emphasize the capacity of the Savage test to detect variance differences more sensitively in small sample sizes.

Analyses performed on larger sample sizes evaluated the performance of the Siegel-Tukey and Savage tests on a larger data set. In the case of equal sample sizes, the type I error rates of the Siegel-Tukey test ranged from 0.047 to 0.049, while the type I error rates of the Savage test ranged from 0.043 to 0.046. These results suggest that both tests have similar type I error rates at large sample sizes, but the Savage test generally offers lower error rates.

In different combinations of sample sizes, the type I error rates of the Siegel-Tukey test ranged between 0.048 and 0.052, while the type I error rates of the Savage test ranged between 0.041 and 0.051. These findings suggest that the Savage test provides lower type I error rates even in large sample sizes and evaluates variance differences more effectively.

It shows that the Savage test is a more reliable option in variance analysis. The low type I error rates in small and large sample sizes suggest that the test is able to detect variance differences more precisely and accurately. Future research based on these findings with larger data sets and different types of distributions will provide a more comprehensive assessment of the overall effectiveness and limitations of the Savage test.

Comparing the obtained results with Tuğran and colleagues' (2015) simulation-based evaluations provides a comprehensive perspective on the performance of the tests. This comparison reveals whether your findings on the performance of Siegel-Tukey and Savage tests are consistent with studies in the literature and offers a deeper understanding of the effectiveness of these tests.

The fact that the Savage test shows lower Type I error rates in small sample sizes, when related to Mukasa and colleagues' (2021) study on the effects of parametric and non-parametric tests, underscores the importance of practical implications in test selection. This correlation helps you better understand the performance of tests in small sample sizes and supports your strategies for obtaining more reliable results in variance analyses.

Furthermore, this study provides important insights for practitioners on how the Savage test can be a more effective tool in dealing with the challenges encountered in analysis of variance. For researchers who want to improve the accuracy of tests and obtain reliable results, especially in small sample sizes, these findings will provide guidance.

The fact that the Savage test shows lower Type I error rates in small sample sizes is based on both theoretical and empirical reasons. Theoretically, the Savage test is a non-parametric test and is less dependent on distributional assumptions, which makes it more robust in small samples. The test's ability to detect variance differences reduces the likelihood of erroneously rejecting the null hypothesis in small sample sizes. Empirically, simulation results and comparative analyses support these characteristics, demonstrating that the test provides lower Type I error rates in small samples. These findings indicate that the Savage test is a reliable and effective analytical tool for small sample sizes.

In conclusion, the findings of this study provide comprehensive information on important factors to consider when evaluating the performance of tests used in analysis of variance. The preference for the Savage test can improve the accuracy and reliability of future studies, thus supporting the quality and validity of research findings in general.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Lowenstein, L.C. (2015) Robustness and Power Comparison of the Mood-Westenberg and Siegel-Tukey Tests.
- [2] Olejnik, S.F. and Algina, J. (1987) Type I Error Rates and Power Estimates of Selected Parametric and Nonparametric Tests of Scale. *Journal of Educational Statistics*, **12**, 45-61. <https://doi.org/10.3102/10769986012001045>
- [3] Tuğran, E., Kocak, M., Mirtağoğlu, H., Yiğit, S. and Mendes, M. (2015) A Simulation Based Comparison of Correlation Coefficients with Regard to Type I Error Rate and Power. *Journal of Data Analysis and Information Processing*, **3**, 87-101. <https://doi.org/10.4236/jdaip.2015.33010>

- [4] Mukasa, E.S., Christospher, W., Ivan, B. and Kizito, M. (2021) The Effects of Parametric, Non-Parametric Tests and Processes in Inferential Statistics for Business Decision Making—A Case of 7 Selected Small Business Enterprises in Uganda. *Open Journal of Business and Management*, **9**, 1510-1526. <https://doi.org/10.4236/ojbm.2021.93081>
- [5] Rousson, V. (2002) On Distribution-Free Tests for the Multivariate Two-Sample Location-Scale Model. *Journal of Multivariate Analysis*, **80**, 43-57. <https://doi.org/10.1006/jmva.2000.1981>
- [6] Lemeshko, B.Y., Lemeshko, S.B. and Gorbunova, A.A. (2010) Application and Power of Criteria for Testing the Homogeneity of Variances. Part I. Parametric Criteria. *Measurement Techniques*, **53**, 237-246. <https://doi.org/10.1007/s11018-010-9489-7>
- [7] Bhattacharyya, H.T. (1977) Nonparametric Estimation of Ratio of Scale Parameters. *Journal of the American Statistical Association*, **72**, 459-463. <https://doi.org/10.1080/01621459.1977.10481021>
- [8] Chenouri, S., Small, C.G. and Farrar, T.J. (2011) Data Depth-Based Nonparametric Scale Tests. *Canadian Journal of Statistics*, **39**, 356-369. <https://doi.org/10.1002/cjs.10099>
- [9] Ehsanes Saleh, A.K. and Dionne, J.-P. (1977) On a Further Generalization of the Savage Test. *Communications in Statistics—Theory and Methods*, **6**, 1213-1221. <https://doi.org/10.1080/03610927708827564>
- [10] Fahoome, G. and Sawilowsky, S.S. (2000) Review of Twenty Nonparametric Statistics and Their Large Sample Approximations.
- [11] Kueppers, S., Rau, R. and Scharf, F. (2024) Using Monte Carlo Simulation to Forecast the Scientific Utility of Psychological App Studies: A Tutorial. *Multivariate Behavioral Research*, **59**, 879-893.
- [12] Abidovna, A.S. (2023) Monte Carlo Modeling and Its Peculiarities in the Implementation of Marketing Analysis in the Activities of the Enterprise. *Gospodarka I Innowacje*, **42**, 375-380.
- [13] Oszczypała, M., Ziółkowski, J. and Małachowski, J. (2023) Modelling the Operation Process of Light Utility Vehicles in Transport Systems Using Monte Carlo Simulation and Semi-Markov Approach. *Energies*, **16**, Article 2210. <https://doi.org/10.3390/en16052210>
- [14] Fleishman, A.I. (1978) A Method for Simulating Non-Normal Distributions. *Psychometrika*, **43**, 521-532. <https://doi.org/10.1007/bf02293811>
- [15] Ramazanov, S. and Senger, Ö. (2023) Küçük, eşit ve büyük örnek hacimlerinde Wald-Wolfowitz, Kolmogorov-Smirnov ve Mann-Whitney testlerinin I. tip hata oranlarının karşılaştırılması. *Ardahan Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, **5**, 137-144. <https://doi.org/10.58588/aru-jfeas.1388231>
- [16] Lee, C.H. (2007) A Monte Carlo Study of Two Nonparametric Statistics with Comparisons of Type I Error Rates and Power. Doctoral Thesis, Oklahoma State University.
- [17] Senger, Ö. (2011) Mann-Whitney, Kolmogorov-Smirnov ve Wald-Wolfowitz testlerinin I. tip hata oranları ve istatistiksel güçleri açısından Monte Carlo Simülasyon çalışması ile karşılaştırılması. Yayınlanmış Doktora Tezi. Erzurum: Atatürk Üniversitesi Sosyal Bilimler Enstitüsü.