Journal of Statistics & Management Systems
ISSN 0972-0510 (Print), ISSN 2169-0014 (Online)
Vol. 28 (2025), No. 3, pp. 559–577
DOI : 10.47974/JSMS-1405

Likelihood-based methods for testing the equality of means of several log-normal populations

Mehmet Cakmak [‡]
Fikri Gökpınar [†]
Meral Ebegil [§]
Esra Gökpınar *
Department of Statistics
Gazi University
Ankara 06500
Turkey

Abstract

In this study, to test the equality of means of several log-normal distributions, our aim is to investigate the effectiveness of likelihood-based methods, such as the likelihood ratio test, the Score test, and the Wald test. For this reason, we propose a Wald test based on the likelihood method. Additionally, the chi-square approximation of the Wald test statistic is often inadequate in small sample scenarios for various problems. To address this, we employ a computational approach test (CAT), known as a type of parametric bootstrap method. To assess the performance of these proposed tests, we evaluate them via their sizes and power under different numbers of groups, sample sizes, and scale parameters. Our simulation study reveals that the CAT of Wald test performs better than the others when the sample sizes and scale parameters within each group are directly proportional. We also provide a data example to illustrate the above methods.

Subject Classification: *62F03, 62F40, 65C05.*

Keywords: *Log-normal distribution, Likelihood-based methods, Parametric bootstrap, Heteroscedasticity, Simulation.*

[‡] *ORCID-ID:* **0000-0001-6036-8956**
[†] *ORCID-ID:* **0000-0002-6310-8727**
[§] *ORCID-ID:* **0000-0003-4798-3422**
* *ORCID-ID:* **0000-0003-2148-4940**

[‡] *E-mail:* **Stat.mehmetcakmak@gmail.com**
[†] *E-mail:* **fikri@gazi.edu.tr**
[§] *E-mail:* **mdemirel@gazi.edu.tr**
* *E-mail:* **eyigit@gazi.edu.tr (Corresponding Author)**

© Taru Publications

1. Introduction

The log-normal distribution is the most common distribution used to model positive and right-skewed data. Many fields including geology and mining, biomedical research, the environment, economics, and others can benefit from this distribution [11]. For example, Lee [9] shows that the log-normal distribution can be used to model the growth rates of soft tissue metastases from breast cancer. Based on the log-normal distribution, Boributh et al. [1] suggested analytical methods to calculate the average wetting ratio in the gas-liquid membrane contacting process. Similarly, Lubrano and Ndoye [13] applied the log-normal distribution to fit the income distribution in the UK.

Since it has a wide range of applications, many theoretical inferences regarding the log-normal distribution have also been made in the literature [17]. A frequent challenge with the log-normal distribution is testing whether several means are equal. While the ANOVA F-test can be applied to assess mean equality under the normality assumption, it is not appropriate when the distribution deviates from normality [8], [10].

Zhou and Tu [18] introduced a likelihood-based test to evaluate the equality of means under lognormal distributions. Gill [6] suggested a modification of the likelihood ratio test (LRT) for this problem; however, as discussed later, this test does not apply to all samples. Wu et al. [16] developed methods using the likelihood method, known as the signed likelihood ratio method (SLRT) and modified SLRT (MSLRT). Lin [12] later introduced a higher-order likelihood technique, also known as the MSLRT, to create a confidence interval of the common mean under the lognormal distributions.

The Wald test (WT), Score test (ST), and LRT are known as the Holy Trinity in asymptotic statistics as described in Rayner's paper [15]. All these tests have good properties in large samples and are asymptotically optimal. Although equivalent in large samples, in small samples the results may be different, and this difference could be potentially important. Although it is commonly believed that asymptotic tests don't work well with small sample sizes, many real-world applications involve samples that are quite small.

Therefore, selecting the appropriate test statistic is crucial. As noted by Rayner [15], if all three tests are easy to compute, then it makes sense to evaluate all three. We have found that our literature does not lead us to choose any of them. So, it is therefore not entirely clear which test should be used in which situation.

As observed in the tests mentioned above, the initial approach often involves using the LRT for testing problems, with significant efforts dedicated to improving the chi-squared approximation of its distribution. In our review of the literature study, we also found that such improvements have not been made much for other tests.

Especially asymptotic tests that converge to the same distribution yield similar results for large sample sizes. However, for small sample sizes, these tests may produce different results due to variations in their exact distributions. Consequently, their use is often unsatisfactory, particularly in many real-world

situations, such as survival analysis, where sample sizes may be relatively small. To address these issues, various resampling techniques have been developed, such as the bootstrap method, and the computational approach test (CAT). Additionally, Davison and Hinkley [4] noted that likelihood-based tests, including the ST, WT, and LRT, are particularly well-suited as test statistics for parametric bootstrap methods. Çakmak et al. [2] explored the ST and its CAT approach.

In the literature review, we have seen that the WT is not investigated to assess mean equality under the log-normal distributions. Furthermore, we have not seen a study in the literature comparing these three tests, namely LRT, WT, and ST together. For this reason, this paper seeks to expand the available methods by introducing the WT to test the equality of means among several log-normal distributions. We have explored the CAT approach of WT to enhance the chi-square approximation of the distribution associated with this test. Moreover, the CAT approaches of ST and WT as well as the CAT approach of LRT were obtained, and these three tests were compared by simulation study.

In light of the objectives outlined above, the remainder of this manuscript is structured as follows: Section 2 describes the LRT, the ST, and their CAT approaches. We also derive the test statistic using the Wald statistic and its CAT approach. Section 3 provides simulation studies to assess the performance of these tests across various scenarios. Section 4 provides an example and discusses the findings of the analysis. At the end, the manuscript concludes with some remarks.

2. Inference on the means of log-normal distributions

To test of equality of means of two and more (k groups) log-normal distributed populations, the relevant inferences are as follows.

$Y_{ij}, i = 1. \dots . k, j = 1, \dots, n_i$ denote a random sample from a log-normal distribution with parameters μ_i and σ_i^2 that is $Y_{ij} \sim LN(\mu_i, \sigma_i^2)$. We express the mean of the ith log normal as $exp(\mu_i + \sigma_i^2/2)$ and we aim to evaluate the equality of these means which is given below:

$$H_0 = exp(\mu_1 + \sigma_1^2/2) = \cdots = exp(\mu_k + \sigma_k^2/2). \quad (1)$$

The logarithms of Y_{ij} are normally distributed as shown $X_{ij} = Ln(Y_{ij}) \sim N(\mu_i, \sigma_i^2)$. Thus, testing H_0 corresponds to testing H_0' as presented below

$$H_0' = \theta_1 = \theta_2 = \cdots = \theta_k, \quad (2)$$

where $\theta_i = \mu_i + \sigma_i^2/2, i = 1. \dots . k$.

Since X_{ij} follows a normal distribution with a mean of $\mu_i = \theta_i - \sigma_i^2/2$ and a variance of σ_i^2, i.e. $X_{ij} \sim N\left(\theta_i - \frac{\sigma_i^2}{2}, \sigma_i^2\right)$. The log-likelihood function is defined as follows

$$l_1 = -\frac{1}{2}\sum_{i=1}^{k} n_i\, ln(2\pi) - \frac{1}{2}\sum_{i=1}^{k} n_i\, ln(\sigma_i^2) - \frac{1}{2}\sum_{i=1}^{k} \sigma_i^{-2} \sum_{j=1}^{n_i}\left(X_{ij} - \left(\theta_i - \frac{\sigma_i^2}{2}\right)\right)^2, \tag{3}$$

For the parameters θ_{i} and σ_{i}^2, the maximum likelihood estimates (MLEs) are obtained as follows.

$$\hat{\theta}_i = \bar{X}_i + S_i^2/2 \text{ and } \hat{\sigma}_i^2 = \frac{1}{n_i}\sum_{j=1}^{n_i}(X_{ij} - \bar{X}_i)^2 = S_i^2, \tag{4}$$

where $\bar{X}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} X_{ij}$, $i = 1. \ldots . k$.

Under $H_0' : \theta_1 = \theta_2 = \cdots = \theta_k = \theta$, the log-likelihood function is obtained as

$$l_0 = -\frac{1}{2}\sum_{i=1}^{k} n_i\, ln(2\pi) - \frac{1}{2}\sum_{i=1}^{k} n_i\, ln(\sigma_i^2) - \frac{1}{2}\sum_{i=1}^{k} \sigma_i^{-2} \sum_{j=1}^{n_i}\left(X_{ij} - \left(\theta - \frac{\sigma_i^2}{2}\right)\right)^2 \tag{5}$$

The restricted maximum likelihood estimates (RMLEs) of θ and σ_i^2 for $i = 1. \ldots . k$ are calculated as

$$\tilde{\theta} = \frac{\sum_{i=1}^{k} n_i\left(\bar{X}_i + \frac{\tilde{\sigma}_i^2}{2}\right)/\tilde{\sigma}_i^2}{\sum_{i=1}^{k}\frac{n_i}{\tilde{\sigma}_i^2}}, \tilde{\sigma}_i^2 = -2 + 2\left(1 + \frac{1}{n_i}\sum_{i=1}^{k}\left(X_{ij} - \tilde{\theta}\right)^2\right)^{\frac{1}{2}} \tag{6}$$

(also seen in [2]). As seen in Equation (6), the RMLEs of θ and σ_i^2do not have an exact form and must be obtained numerically. The equations presented above can be found iteratively by applying the below algorithm.

(1) Set $\sigma_{i,0}^2 = S_i^2$ for $i = 1, \ldots, k$.

(2) Obtain $\tilde{\theta}_{h+1} = \dfrac{\sum_{i=1}^{k} n_i\left(\bar{X}_i + \frac{\tilde{\sigma}_{i,h}^2}{2}\right)\Big/\tilde{\sigma}_{i,h}^2}{\sum_{i=1}^{k}\frac{n_i}{\tilde{\sigma}_{i,h}^2}}$ and

$$\tilde{\sigma}_{i,h+1}^2 = -2 + 2\left(1 + \frac{1}{n_i}\sum_{i=1}^{k}(X_{ih} - \tilde{\theta}_{h+1})^2\right)^{1/2} \text{ for } h = 0,1,2, \ldots$$

(3) If $\left|\tilde{\theta}_{h+1} - \tilde{\theta}_h\right| <$ error tolerance. stop.

(4) Else set $\tilde{\theta}_h = \tilde{\theta}_{h+1}$.

(5) Repeat steps 2 to 4 until the desired error tolerance is reached.

The simulation study reveals that this method, known as the fixed-point approach, converges rapidly in about 10 iterations.

2.1 *Likelihood Ratio Test*

To test H_0' in Equation (2), the LRT is based on the statistic

$$\Lambda = 2\left(l_1\left(\hat{\theta}_1, \ldots, \hat{\theta}_k, \hat{\sigma}_1^2, \ldots, \hat{\sigma}_k^2\right) - l_0(\tilde{\theta}, \tilde{\sigma}_1^2, \ldots, \tilde{\sigma}_k^2)\right)$$

By using the estimates defined in Equation (4) and Equation (6), the likelihood ratio test given can be expressed as follows:

$$\Lambda = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\frac{(X_{ij}-\tilde{\theta}+\frac{\tilde{\sigma}_i^2}{2})^2}{\tilde{\sigma}_i^2} + \sum_{i=1}^{k} n_i \, ln\left(\frac{n_i\tilde{\sigma}_i^2}{(n_i-1)S_i^2}\right) - \sum_{i=1}^{k} n_i \,, \tag{7}$$

where $\Lambda \sim \chi_{k-1}^2$ under H_0'. Λ rejects H_0' at level α if $\Lambda > \chi_{k-1,\alpha}^2$, where $\chi_{k-1,\alpha}^2$ express the α th upper quantile of χ_{k-1}^2 [2].

2.2. *Score test*

To describe this test, let $\underline{\theta} = (\theta_1, \theta_2, \dots, \theta_k)$, $\underline{\sigma}^2 = (\sigma_1^2, \sigma_2^2, \dots, \sigma_k^2)$. The general definition of the Score test (ST) statistic given by Çakmak et al. [2] is shown in the Equation (3)

$$S = U'_{\underline{\theta}}\left(\underline{\tilde{\theta}}, \underline{\tilde{\sigma}}^2\right)\left(I^{\underline{\theta},\underline{\theta}}\left(\underline{\tilde{\theta}}, \underline{\tilde{\sigma}}^2\right)\right) U_{\underline{\theta}}\left(\underline{\tilde{\theta}}, \underline{\tilde{\sigma}}^2\right)$$

$\underline{\tilde{\theta}}$ and $\underline{\tilde{\sigma}}^2$ are the RMLEs of $\underline{\theta}$ and $\underline{\sigma}^2$ parameters, respectively. Derived from the log-likelihood function, the Score vector $U'_{\underline{\theta}}\left(\underline{\tilde{\theta}}, \underline{\tilde{\sigma}}^2\right)$ is calculated using RMLEs $\underline{\tilde{\theta}}$ and $\underline{\tilde{\sigma}}^2$. Under the restricted model, $I^{\underline{\theta},\underline{\theta}}\left(\underline{\tilde{\theta}}, \underline{\tilde{\sigma}}^2\right)$ denotes the upper right block matrix of the inverse of the information matrix.

ST is given by

$$S = \sum_{i=1}^{k}\frac{n_i(2+\tilde{\sigma}_i^2)\left(\bar{X}_i-\tilde{\theta}-\frac{\tilde{\sigma}_i^2}{2}\right)^2}{2\tilde{\sigma}_i^2}\,. \tag{8}$$

Under the null hypothesis, $S \sim \chi_{k-1}^2$. If $S > \chi_{k-1,\alpha}^2$, then S rejects H_0' at level α. Here, $\chi_{k-1,\alpha}^2$ demonstrates the α th upper quantile of χ_{k-1}^2.

2.3 *Wald test*

We develop a Wald test to test H_0' in Equation (2). To describe this test, let $\underline{\theta} = (\theta_1, \theta_2, \dots, \theta_k)$, $\underline{\sigma}^2 = (\sigma_1^2, \sigma_2^2, \dots, \sigma_k^2)$. The general definition of the WT statistic is shown in the Equation (9) below by Engle [5]

$$W = \left(\hat{\theta} - \tilde{\theta}\right)' \left(I^{\underline{\theta},\underline{\theta}}\left(\underline{\hat{\theta}}, \underline{\hat{\sigma}}^2\right)\right)^{-1} \left(\hat{\theta} - \tilde{\theta}\right) \tag{9}$$

where $\left(\hat{\theta} - \tilde{\theta}\right)' = \left[\hat{\theta}_1 - \tilde{\theta}, \hat{\theta}_2 - \tilde{\theta}, \dots, \hat{\theta}_k - \tilde{\theta}\right]$. $\hat{\theta}$ is the MLE of $\underline{\theta} = (\theta_1, \theta_2, \dots, \theta_k)$ and $\tilde{\theta}$ is the RMLE of θ.

The MLEs of parameters σ_i^2 are $\sigma_1^2 = \hat{\sigma}_1^2, \sigma_2^2 = \hat{\sigma}_2^2, \dots, \sigma_k^2 = \hat{\sigma}_k^2$ and when the estimates are placed, the result is obtained as

$$I^{\underline{\theta},\underline{\theta}} = \left(I_{\underline{\theta},\underline{\theta}} - I_{\underline{\theta},\underline{\sigma}^2} I^{-1}_{\underline{\sigma}^2,\underline{\sigma}^2} I_{\underline{\sigma}^2,\underline{\theta}}\right)^{-1} = \begin{bmatrix} \frac{\tilde{\sigma}_1^2}{n_i}\left(1+\frac{\tilde{\sigma}_1^2}{2}\right) & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \frac{\tilde{\sigma}_k^2}{n_k}\left(1+\frac{\tilde{\sigma}_k^2}{2}\right) \end{bmatrix}$$

The calculation of the Wald test statistic is illustrated in Equation (10).

$$W = \left(\hat{\theta}_1 - \tilde{\theta}, \hat{\theta}_2 - \tilde{\theta}, \ldots, \hat{\theta}_k - \tilde{\theta}\right) \begin{bmatrix} \frac{\hat{\sigma}_1^2}{n_1}\left(1 + \frac{\hat{\sigma}_1^2}{2}\right) & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & \frac{\hat{\sigma}_k^2}{n_k}\left(1 + \frac{\hat{\sigma}_k^2}{2}\right) \end{bmatrix}^{-1} \begin{pmatrix} \hat{\theta}_1 - \tilde{\theta} \\ \hat{\theta}_2 - \tilde{\theta} \\ \vdots \\ \hat{\theta}_k - \tilde{\theta} \end{pmatrix}$$

$$W = \sum_{i=1}^{k} \left(\frac{2n_i\left(\hat{\theta}_i - \tilde{\theta}\right)^2}{\hat{\sigma}_i^2\left(2 + \hat{\sigma}_i^2\right)} \right) \tag{10}$$

W$\sim\chi^2_{k-1}$ under the null hypothesis.

2.4 *CAT approach*

To gain a better understanding of the CAT, we begin by outlining the general framework of this method. Consider a random sample represented by $U_1, \ldots, U_n$ with a known probability density function $f(u/\theta)$, where the specific form of *f* is taken to be defined. Here $\theta = \left(\theta^{(1)}, \theta^{(2)}\right)$ is an unknown vector within the parameter space Θ. In this context, $\theta^{(1)}$ is the parameter of interest, while $\theta^{(2)}$ is considered a nuisance parameter. The primary focus is on testing the hypothesis $H_0' : \theta^{(1)} = \theta_0^{(1)}$ against an appropriate alternative, H_1'. To evaluate H_0' versus H_1', we initially reformulate H_0' given as $H_0'^{*} : \eta\left(\theta^{(1)}, \theta_0^{(1)}\right) = 0$ against $H_1'^{*}$, where η is a scalar-valued function. The following stages illustrate the typical CAT methodology to test $H_0'^{*}$ at a specified level α [14].

1. Calculate the MLEs of $\Theta^{(1)}$ and $\Theta^{(2)}$.
2. Identify an appropriate function $\eta\left(\theta^{(1)}, \theta_0^{(1)}\right)$, and use the MLEs $\hat{\eta}\left(\hat{\theta}^{(1)}, \theta_0^{(1)}\right)$ as a test statistic.
3. Determine the MLEs for the $\Theta^{(2)}$ parameter, denoted as $\tilde{\theta}^{(2)}$ under H_o.
4. Generate an artificial sample $U_1, \ldots, U_n$ from $f\left(u/\theta_0^{(1)}, \tilde{\theta}^{(2)}\right)$ many times, for example, *m* times. Compute the MLE $\tilde{\eta}^{(j)}$, *j*=1,…,*m* for each of these generated samples.
5. Calculate the *p*-value as $\hat{p} = \#\left(\hat{\eta}^{(j)} > \hat{\eta}\right)/m$. If $\hat{p} < \alpha$, then H_0 is rejected.

RMLEs form the basis of CAT. This method is very simple to use because the *p*-value may be computed directly, and the theoretical distribution of the test statistic does not need to be obtained. The CAT-based testing process for determining whether the means of the several log-normal distributions are equal may therefore be provided as follows:

LRT's CAT Approach

We derive the CAT method using the LRT statistic to test H_0' in Equation (2). The algorithm of this method is as follows:

For a set of *k* samples using log transformation:

(1) Obtain the MLEs $\hat{\theta}_i, \hat{\sigma}_i^2$ and the restricted MLEs $\tilde{\theta}, \tilde{\sigma}_i^2$, $i = 1, \ldots, k$.

(2) Calculate the Λ statistics in Equation (7).
(3) Draw a pseudo random sample with size n_i from the $X_{ij} \sim N\left(\tilde{\theta} - 0.5\tilde{\sigma}_i^2, \tilde{\sigma}_i^2\right)$ for $i = 1, \ldots, k$.
(4) Compute the Λ statistic for the generated samples.
(5) Repeat steps 3 and 4 for many times (for instance, L times). Denote the calculated Λ statistic as $\Lambda^{(l)}$, l=1,…, L for each of these replicated samples.
(6) Calculate the p-value as $\hat{p} = \sum_{l=1}^{L} I\left(\Lambda^{(l)} > \Lambda\right)/L$. Here $I(.)$ represents the indicator function.
(7) Reject the H_0' in Equation (2) if the $\hat{p} \leq \alpha$.

ST's CAT Approach

The algorithm of CAT method based on ST statistic is as follows:

For a set of k samples using log-transformation:
(1) Compute the MLEs $\hat{\theta}_i, \hat{\sigma}_i^2$ and the restricted MLEs $\tilde{\theta}, \tilde{\sigma}_i^2$. $i = 1, \ldots, k$.
(2) Calculate the S test statistics given in Equation (8) by using the observed value.
(3) Draw a pseudo random sample with size n_i from the $X_{ij} \sim N\left(\tilde{\theta} - 0.5\tilde{\sigma}_i^2, \tilde{\sigma}_i^2\right), i = 1, \ldots, k$.
(4) Compute the ST statistic for the generated samples.
(5) Repeat the step 3 and 4 for many times (for instance, L times). Calculate ST statistic, denoted by $\mathrm{ST}^{(l)}$, l=1,…,L for each of these generated samples.
(6) Obtain the estimated p-value as $\hat{p} = \sum_{l=1}^{L} I\left(\mathrm{ST}^{(l)} > \mathrm{ST}\right)/L$ for l=1,…,L.
(7) If the $\hat{p} \leq \alpha$, reject the H_0' in Equation (2).

WT's CAT Approach

We propose to use the CAT method using the Wald test statistic to test H_0' in Equation (2). The algorithm of this method is as follows:

For a set of k samples using log-transformation:
(1) Obtain the MLEs $\hat{\theta}_i, \hat{\sigma}_i^2$ and the restricted MLEs $\tilde{\theta}, \tilde{\sigma}_i^2$. $i = 1, \ldots, k$.
(2) Calculate the W test statistics in Equation (10).
(3) Draw a pseudo random sample with size n_i from the $X_{ij} \sim N\left(\tilde{\theta} - 0.5\tilde{\sigma}_i^2, \tilde{\sigma}_i^2\right), i = 1, \ldots, k$.
(4) Calculate the W test statistic for the generated samples.
(5) Repeat the step 3 and 4 for many times (for instance, L times). Calculate W test statistic, denoted by $W^{(l)}$, l=1,…, L for each of these generated samples.
(6) Compute the estimated p-value as $\hat{p} = \sum_{l=1}^{L} I\left(W^{(l)} > W\right)/L$ for l=1,…,L.
(7) If the $\hat{p} \leq \alpha$, reject the H_0' in Equation (2).

3. Simulation Study

In this part of the study, we assess the effectiveness of the CAT methods related to likelihood-based techniques by examining the estimated sizes and powers of the tests at a nominal level of 0.05. To accomplish this, we perform a simulation study to evaluate the CAT methods of the likelihood-based approaches, specifically the CAT approach of the Wald test (WCAT), the CAT approach of the likelihood ratio test (LRTCAT), and the CAT approach of the Score test (SCAT).

We analyzed all tests under different situations to determine which test performs best in which situation. For this purpose, a simulation study was created under different values of the number of groups and population variances. We analyzed scenarios with equal sample sizes as well as scenarios with unequal sample sizes. We also considered cases where sample sizes are inversely proportional to population variances and cases where population variances are directly proportional to sample sizes. Consequently, the random samples were drawn from a log-normal distribution setting $\mu_i = \theta_i - \sigma_i^2/2$ and σ_i^2. 3000 replications were performed to calculate the sizes and the powers of the tests. Additionally, the p-value estimates for the CAT approaches were derived using Monte Carlo simulations with L=3000. The calculated sizes of all tests are presented in Tables 1 to 3 for α=0.05.

Table 1
For k=3, the calculated sizes of all tests

σ_i^2	n	LRT	WT	ST	LRCAT	WCAT	SCAT
0.1, 0.5, 1	5, 5,5	0.109	0.230	0.058	0.056	0.053	0.055
	10,10,10	0.079	0.135	0.050	0.054	0.048	0.049
	20,20,20	0.065	0.091	0.048	0.053	0.042	0.048
	30,30,30	0.064	0.081	0.050	0.053	0.045	0.050
	3,9,15	0.083	0.150	0.047	0.051	0.042	0.047
	5,15,20	0.072	0.107	0.048	0.054	0.051	0.050
	10,20,30	0.055	0.083	0.054	0.049	0.049	0.055
	15,9,3	0.119	0.273	0.056	0.043	0.051	0.051
	20.15.5	0.095	0.190	0.045	0.060	0.058	0.040
	30,20,10	0.071	0.120	0.053	0.053	0.051	0.055
0.25, 0.75, 1.25	5, 5,5	0.100	0.235	0.054	0.047	0.053	0.050
	10,10,10	0.079	0.143	0.040	0.051	0.046	0.042
	20,20,20	0.057	0.090	0.053	0.047	0.044	0.052
	30,30,30	0.061	0.077	0.048	0.054	0.050	0.050
	3,9,15	0.082	0.170	0.056	0.046	0.050	0.053
	5,15,20	0.070	0.111	0.047	0.049	0.045	0.051
	10,20,30	0.061	0.087	0.050	0.048	0.044	0.051
	15,9,3	0.139	0.274	0.059	0.059	0.049	0.058
	20,15,5	0.093	0.179	0.045	0.052	0.056	0.039
	30,20,10	0.060	0.115	0.051	0.044	0.044	0.049
0.5, 1.25, 2	5, 5,5	0.096	0.253	0.052	0.042	0.054	0.048

	10,10,10	0.082	0.157	0.047	0.057	0.049	0.049
	20,20,20	0.061	0.099	0.048	0.051	0.051	0.048
	30,30,30	0.061	0.081	0.052	0.053	0.049	0.054
	3,9,15	0.093	0.183	0.059	0.045	0.052	0.057
	5,15,20	0.080	0.138	0.053	0.053	0.049	0.054
	10,20,30	0.064	0.100	0.051	0.051	0.048	0.052
	15,9,3	0.126	0.302	0.050	0.052	0.060	0.043
	20.15.5	0.087	0.194	0.039	0.054	0.054	0.038
	30,20,10	0.070	0.119	0.058	0.050	0.049	0.057

Table 2
For $k = 4$, the calculated sizes of all tests

σ_i^2	n	LRT	WT	ST	LRCAT	WCAT	SCAT
0.1, 0.5,	5, 5,5,5	0.119	0.289	0.057	0.058	0.053	0.050
0.5, 1	10,10,10,10	0.085	0.169	0.049	0.050	0.058	0.046
	20,20,20,20	0.068	0.110	0.056	0.055	0.051	0.056
	30,30,30,30	0.063	0.089	0.059	0.051	0.050	0.056
	3,9,9,15	0.091	0.180	0.051	0.047	0.056	0.052
	5,15,15,20	0.079	0.139	0.048	0.052	0.048	0.048
	10,20,20,30	0.066	0.107	0.048	0.053	0.051	0.049
	15,9,9,3	0.129	0.278	0.060	0.046	0.050	0.054
	20,15,15,5	0.088	0.201	0.056	0.053	0.052	0.052
	30,20,20,10	0.080	0.134	0.058	0.055	0.048	0.055
0.25, 0.75,	5, 5,5,5	0.119	0.291	0.060	0.047	0.046	0.052
0.75, 1.25	10,10,10,10	0.083	0.173	0.056	0.049	0.049	0.054
	20,20,20,20	0.064	0.101	0.053	0.049	0.057	0.055
	30,30,30,30	0.064	0.092	0.050	0.054	0.045	0.050
	3,9,9,15	0.107	0.228	0.051	0.056	0.052	0.046
	5,15,15,20	0.080	0.146	0.051	0.058	0.059	0.053
	10,20,20,30	0.061	0.103	0.051	0.047	0.051	0.050
	15,9,9,3	0.142	0.324	0.054	0.059	0.056	0.050
	20,15,15,5	0.084	0.187	0.058	0.049	0.057	0.058
	30,20,20,10	0.077	0.133	0.051	0.061	0.045	0.050
0.5, 1.25,	5, 5,5,5	0.111	0.307	0.060	0.049	0.050	0.051
1.25, 2	10,10,10,10	0.080	0.189	0.054	0.052	0.046	0.051
	20,20,20,20	0.060	0.109	0.054	0.045	0.053	0.054
	30,30,30,30	0.061	0.092	0.047	0.052	0.046	0.047
	3,9,9,15	0.099	0.242	0.060	0.049	0.050	0.054
	5,15,15,20	0.078	0.165	0.059	0.052	0.057	0.058
	10,20,20,30	0.059	0.108	0.058	0.050	0.055	0.058
	15,9,9,3	0.127	0.329	0.056	0.052	0.051	0.043
	20,15,15,5	0.095	0.220	0.057	0.049	0.043	0.054
	30,20,20,10	0.067	0.145	0.050	0.049	0.050	0.046

Table 3
For k = 5, the calculated sizes of all tests

σ_i^2	n	LRT	WT	ST	LRCAT	WCAT	SCAT
0.1, 0.1, 0.5, 1, 1	5,5,5,5,5	0.143	0.339	0.044	0.057	0.040	0.042
	10,10,10,10,10	0.092	0.188	0.056	0.055	0.058	0.055
	20,20,20,20,20	0.066	0.116	0.045	0.050	0.047	0.042
	30,30,30,30,30	0.066	0.099	0.050	0.055	0.044	0.049
	3,9,9,15,15	0.101	0.232	0.051	0.051	0.031	0.050
	5,15,15,20,20	0.078	0.152	0.047	0.050	0.046	0.049
	10,20,20,30,30	0.064	0.105	0.045	0.048	0.040	0.048
	15,15,9,9,3	0.122	0.307	0.060	0.053	0.047	0.057
	20,20,15,15,5	0.099	0.249	0.065	0.045	0.065	0.059
	30,30,20,20,10	0.068	0.138	0.053	0.053	0.054	0.047
0.25, 0.25, 0.75, 1.25, 1.25	5,5,5,5,5	0.138	0.360	0.060	0.055	0.036	0.051
	10,10,10,10,10	0.095	0.198	0.057	0.055	0.047	0.050
	20,20,20,20,20	0.060	0.110	0.050	0.048	0.044	0.048
	30,30,30,30,30	0.060	0.098	0.052	0.055	0.050	0.050
	3,9,9,15,15	0.106	0.251	0.050	0.051	0.048	0.045
	5,15,15,20,20	0.083	0.159	0.045	0.054	0.051	0.045
	10,20,20,30,30	0.075	0.124	0.051	0.056	0.053	0.050
	15,15,9,9,3	0.120	0.312	0.065	0.044	0.059	0.060
	20,20,15,15,5	0.105	0.272	0.060	0.052	0.048	0.053
	30,30,20,20,10	0.070	0.122	0.059	0.046	0.054	0.056
0.5, 0.5, 1.25, 2, 2	5,5,5,5,5	0.131	0.369	0.059	0.045	0.044	0.058
	10,10,10,10,10	0.092	0.222	0.055	0.058	0.045	0.049
	20,20,20,20,20	0.061	0.127	0.048	0.048	0.041	0.048
	30,30,30,30,30	0.060	0.104	0.046	0.053	0.043	0.043
	3,9,9,15,15	0.107	0.257	0.059	0.050	0.043	0.046
	5,15,15,20,20	0.089	0.183	0.049	0.055	0.051	0.047
	10,20,20,30,30	0.067	0.127	0.050	0.050	0.051	0.050
	15,15,9,9,3	0.125	0.352	0.067	0.057	0.060	0.055
	20,20,15,15,5	0.107	0.285	0.056	0.054	0.054	0.046
	30,30,20,20,10	0.067	0.151	0.051	0.048	0.051	0.044

Tables 1 to 3 indicate that the sizes of the LRT and WT are significantly higher than the nominal level, with the WT exhibiting the highest values. In contrast, the sizes of the ST are very close to the nominal level. However, in cases where the number of groups is large and the sample size is small, the size of the ST is slightly above the nominal level.

As can be seen, the sizes of the WT and LRT are higher than those of other tests. In addition, the sizes of the WCAT, LRCAT, and SCAT are very close to the nominal level in all considered cases. Given that the LRT, WT, and ST are asymptotic tests that converge to a chi-square distribution, their sizes are likely to exceed the nominal level. Nevertheless, the results indicate that the ST and SCAT tests indeed have sizes close to the nominal level, which is an important and noteworthy finding.

For the purpose of assessing the power of all tests, for i=1,...,k we considered various values of θ_i's and σ_i^2 . For α=0.05, the calculated powers of all tests are given in Tables 4 to 9.

Table 4
For k = 3 and θ = (0.1, 0.5, 1), the calculated powers of all tests

σ_i^2	n	LRT	WT	ST	LRCAT	WCAT	SCAT
0.1, 0.5, 1	5, 5,5	0.485	0.503	0.400	0.332	0.156	0.401
	10,10,10	0.737	0.690	0.753	0.676	0.428	0.761
	20,20,20	0.954	0.939	0.971	0.945	0.899	0.971
	30,30,30	0.995	0.991	0.997	0.993	0.987	0.997
	3,9,15	0.624	0.782	0.288	0.449	0.445	0.293
	5,15,20	0.828	0.912	0.483	0.755	0.785	0.515
	10,20,30	0.961	0.972	0.905	0.949	0.941	0.909
	15,9,3	0.524	0.481	0.600	0.331	0.059	0.597
	20.15.5	0.707	0.619	0.769	0.603	0.166	0.764
	30,20,10	0.883	0.832	0.907	0.853	0.595	0.912
0.25, 0.75, 1.25	5, 5,5	0.333	0.403	0.233	0.205	0.119	0.225
	10,10,10	0.510	0.524	0.489	0.438	0.284	0.496
	20,20,20	0.814	0.800	0.813	0.794	0.680	0.815
	30,30,30	0.941	0.929	0.940	0.929	0.885	0.941
	3,9,15	0.411	0.639	0.164	0.256	0.266	0.154
	5,15,20	0.565	0.742	0.294	0.465	0.510	0.295
	10,20,30	0.770	0.838	0.616	0.724	0.742	0.626
	15,9,3	0.379	0.368	0.439	0.218	0.052	0.420
	20,15,5	0.523	0.435	0.594	0.424	0.088	0.584
	30,20,10	0.691	0.603	0.756	0.646	0.352	0.758
0.5, 1.25, 2	5, 5,5	0.209	0.311	0.167	0.115	0.074	0.158
	10,10,10	0.286	0.317	0.260	0.233	0.131	0.268
	20,20,20	0.498	0.479	0.505	0.469	0.319	0.509
	30,30,30	0.666	0.635	0.684	0.642	0.559	0.686
	3,9,15	0.267	0.477	0.097	0.156	0.155	0.089
	5,15,20	0.317	0.505	0.161	0.245	0.279	0.159
	10,20,30	0.440	0.552	0.331	0.397	0.426	0.337
	15,9,3	0.244	0.270	0.298	0.125	0.041	0.278
	20,15,5	0.306	0.225	0.373	0.219	0.038	0.364
	30,20,10	0.413	0.318	0.457	0.363	0.114	0.455

Table 5
For k = 3 and θ = (0.25, 0.75, 1.25), the calculated powers of all tests

σ_i^2	n	LRT	WT	ST	LRCAT	WCAT	SCAT
0.1, 0.5, 1	5, 5,5	0.576	0.589	0.418	0.412	0.224	0.419
	10,10,10	0.859	0.831	0.850	0.817	0.604	0.855
	20,20,20	0.986	0.981	0.990	0.981	0.962	0.991
	30,30,30	1.000	1.000	1.000	1.000	0.998	1.000
	3,9,15	0.674	0.858	0.273	0.485	0.536	0.261
	5,15,20	0.892	0.962	0.488	0.824	0.885	0.491

	10,20,30	0.983	0.989	0.938	0.979	0.984	0.942
	15,9,3	0.669	0.615	0.737	0.478	0.081	0.725
	20.15.5	0.844	0.783	0.884	0.777	0.265	0.880
	30,20,10	0.952	0.926	0.976	0.935	0.798	0.977
	5, 5,5	0.393	0.466	0.266	0.248	0.149	0.258
	10,10,10	0.618	0.642	0.582	0.548	0.436	0.595
	20,20,20	0.892	0.890	0.895	0.873	0.819	0.897
	30,30,30	0.972	0.970	0.982	0.970	0.958	0.982
0.25, 0.75,	3,9,15	0.460	0.696	0.151	0.295	0.293	0.137
1.25	5,15,20	0.632	0.810	0.286	0.521	0.597	0.284
	10,20,30	0.838	0.908	0.659	0.803	0.829	0.664
	15,9,3	0.490	0.448	0.548	0.294	0.069	0.530
	20.15.5	0.648	0.572	0.700	0.554	0.154	0.696
	30,20,10	0.822	0.757	0.858	0.785	0.540	0.858
	5, 5,5	0.231	0.340	0.164	0.125	0.088	0.153
	10,10,10	0.334	0.390	0.301	0.278	0.181	0.305
	20,20,20	0.584	0.587	0.607	0.551	0.466	0.613
	30,30,30	0.768	0.760	0.793	0.754	0.695	0.795
0.5, 1.25,	3,9,15	0.300	0.522	0.106	0.174	0.178	0.096
2	5,15,20	0.375	0.578	0.155	0.283	0.318	0.154
	10,20,30	0.510	0.648	0.371	0.456	0.515	0.377
	15,9,3	0.301	0.306	0.342	0.164	0.047	0.317
	20.15.5	0.374	0.296	0.454	0.273	0.055	0.444
	30,20,10	0.520	0.431	0.575	0.470	0.188	0.576

Table 6
For $k = 4$ and $\theta = (0.1, 0.5, 0.5, 1)$, the calculated powers of all tests

σ_i^2	n	LRT	WT	ST	LRCAT	WCAT	SCAT
	5, 5,5,5	0.464	0.533	0.354	0.272	0.161	0.344
	10,10,10,10	0.727	0.721	0.691	0.645	0.473	0.701
	20,20,20,20	0.954	0.944	0.962	0.945	0.896	0.963
	30,30,30,30	0.996	0.994	0.997	0.995	0.989	0.997
0.1, 0.5,	3,9,9,15	0.561	0.763	0.311	0.404	0.397	0.309
0.5,1	5,15,15,20	0.753	0.879	0.515	0.675	0.704	0.515
	10,20,20,30	0.936	0.960	0.832	0.915	0.918	0.833
	15,9,9,3	0.599	0.567	0.659	0.399	0.082	0.651
	20,15,15,5	0.793	0.737	0.843	0.713	0.260	0.836
	30,20,20,10	0.919	0.888	0.953	0.896	0.717	0.953
	5, 5,5,5	0.310	0.437	0.242	0.165	0.100	0.225
	10,10,10,10	0.479	0.529	0.430	0.397	0.269	0.429
	20,20,20,20	0.762	0.775	0.763	0.732	0.645	0.764
	30,30,30,30	0.917	0.917	0.921	0.907	0.868	0.923
0.25, 0.75,	3,9,9,15	0.380	0.616	0.196	0.256	0.246	0.183
0.75, 1.25	5,15,15,20	0.518	0.704	0.316	0.415	0.443	0.309
	10,20,20,30	0.709	0.803	0.574	0.654	0.656	0.577
	15,9,9,3	0.405	0.427	0.446	0.227	0.048	0.420
	20,15,15,5	0.533	0.479	0.613	0.433	0.127	0.602
	30,20,20,10	0.708	0.659	0.750	0.662	0.409	0.749

	5, 5,5,5	0.217	0.372	0.159	0.104	0.060	0.138
	10,10,10,10	0.279	0.355	0.252	0.205	0.119	0.248
	20,20,20,20	0.435	0.453	0.451	0.400	0.293	0.449
	30,30,30,30	0.594	0.602	0.617	0.571	0.473	0.619
0.5, 1.25,	3,9,9,15	0.251	0.462	0.117	0.143	0.139	0.109
1.25, 2	5,15,15,20	0.295	0.505	0.159	0.216	0.234	0.167
	10,20,20,30	0.384	0.504	0.288	0.340	0.343	0.284
	15,9,9,3	0.244	0.295	0.294	0.126	0.036	0.263
	20,15,15,5	0.313	0.281	0.377	0.226	0.039	0.363
	30,20,20,10	0.398	0.353	0.447	0.353	0.136	0.441

Table 7
For $k = 4$ and $\theta = (0.25, 0.75, 0.75, 1.25)$, the calculated powers of all tests

σ_i^2	n	LRT	WT	ST	LRCAT	WCAT	SCAT
	5, 5,5,5	0.577	0.671	0.374	0.368	0.242	0.370
	10,10,10,10	0.845	0.866	0.738	0.783	0.654	0.741
	20,20,20,20	0.993	0.993	0.990	0.992	0.972	0.990
	30,30,30,30	0.999	0.999	1.000	0.999	0.998	1.000
0.1, 0.5,	3,9,9,15	0.626	0.851	0.291	0.437	0.463	0.282
0.5,1	5,15,15,20	0.842	0.947	0.510	0.764	0.818	0.510
	10,20,20,30	0.972	0.988	0.848	0.959	0.976	0.848
	15,9,9,3	0.758	0.730	0.797	0.589	0.134	0.783
	20,15,15,5	0.918	0.885	0.940	0.873	0.463	0.936
	30,20,20,10	0.983	0.975	0.992	0.975	0.898	0.992
	5, 5,5,5	0.364	0.517	0.252	0.197	0.138	0.233
	10,10,10,10	0.577	0.645	0.466	0.488	0.377	0.461
	20,20,20,20	0.875	0.896	0.845	0.857	0.793	0.846
	30,30,30,30	0.972	0.976	0.968	0.968	0.960	0.967
0.25, 0.75,	3,9,9,15	0.425	0.692	0.186	0.259	0.265	0.178
0.75, 1.25	5,15,15,20	0.571	0.777	0.325	0.455	0.503	0.315
	10,20,20,30	0.793	0.890	0.573	0.748	0.774	0.575
	15,9,9,3	0.526	0.543	0.512	0.339	0.067	0.491
	20,15,15,5	0.711	0.672	0.724	0.611	0.227	0.719
	30,20,20,10	0.848	0.818	0.870	0.809	0.629	0.867
	5, 5,5,5	0.225	0.403	0.151	0.114	0.071	0.150
	10,10,10,10	0.312	0.412	0.279	0.240	0.160	0.273
	20,20,20,20	0.552	0.594	0.532	0.513	0.431	0.532
	30,30,30,30	0.729	0.752	0.710	0.705	0.630	0.712
0.5, 1.25,	3,9,9,15	0.271	0.535	0.126	0.151	0.171	0.107
1.25, 2	5,15,15,20	0.331	0.553	0.172	0.245	0.275	0.166
	10,20,20,30	0.450	0.603	0.313	0.404	0.432	0.314
	15,9,9,3	0.309	0.364	0.328	0.171	0.048	0.295
	20,15,15,5	0.387	0.367	0.461	0.287	0.084	0.449
	30,20,20,10	0.520	0.481	0.573	0.465	0.226	0.570

Table 8
For k = 5 and θ = (0.1, 0.1, 0.5, 1, 1), the calculated powers of all tests

σ_i^2	n	LR	W	S	LRCAT	WCAT	SCAT
0.1, 0.1, 0.5, 1,1	5,5,5,5,5	0.672	0.666	0.669	0.475	0.143	0.663
	10,10,10,10,10	0.894	0.847	0.919	0.847	0.520	0.918
	20,20,20,20,20	0.998	0.993	0.999	0.997	0.978	0.999
	30,30,30,30,30	1.000	1.000	1.000	1.000	1.000	1.000
	3,9,9,15,15	0.863	0.948	0.380	0.656	0.536	0.347
	5,15,15,20,20	0.972	0.988	0.699	0.936	0.928	0.688
	10,20,20,30,30	0.997	0.999	0.996	0.997	0.993	0.996
	15,15,9,9,3	0.648	0.622	0.713	0.395	0.039	0.702
	20,20,15,15,5	0.811	0.715	0.883	0.713	0.108	0.873
	30,30,20,20,10	0.963	0.916	0.979	0.946	0.665	0.980
0.25,0.25, 0.75, 1.25, 1.25	5,5,5,5,5	0.481	0.560	0.420	0.277	0.111	0.392
	10,10,10,10,10	0.724	0.713	0.737	0.634	0.370	0.729
	20,20,20,20,20	0.952	0.940	0.964	0.943	0.856	0.963
	30,30,30,30,30	0.996	0.995	0.996	0.995	0.984	0.996
	3,9,9,15,15	0.607	0.837	0.218	0.364	0.297	0.188
	5,15,15,20,20	0.801	0.921	0.421	0.682	0.679	0.400
	10,20,20,30,30	0.945	0.972	0.862	0.921	0.917	0.860
	15,15,9,9,3	0.502	0.496	0.581	0.267	0.037	0.544
	20,20,15,15,5	0.655	0.537	0.743	0.526	0.055	0.727
	30,30,20,20,10	0.844	0.748	0.900	0.807	0.379	0.894
0.5, 0.5, 1.25, 2, 2	5,5,5,5,5	0.299	0.448	0.280	0.146	0.060	0.237
	10,10,10,10,10	0.409	0.447	0.446	0.316	0.149	0.430
	20,20,20,20,20	0.678	0.654	0.717	0.639	0.440	0.711
	30,30,30,30,30	0.855	0.835	0.880	0.840	0.715	0.875
	3,9,9,15,15	0.367	0.683	0.147	0.185	0.162	0.116
	5,15,15,20,20	0.488	0.726	0.214	0.349	0.354	0.191
	10,20,20,30,30	0.664	0.788	0.521	0.595	0.587	0.510
	15,15,9,9,3	0.321	0.375	0.385	0.139	0.027	0.343
	20,20,15,15,5	0.369	0.290	0.487	0.250	0.020	0.457
	30,30,20,20,10	0.529	0.380	0.644	0.465	0.092	0.632

Table 9
For k = 5 and θ = (0.25, 0.25, 0.75, 1.25, 1.25), the calculated powers of all tests

σ_i^2	n	LRT	WT	ST	LRCAT	WCAT	SCAT
0.1, 0.1, 0.5, 1, 1	5,5,5,5,5	0.766	0.761	0.723	0.576	0.198	0.714
	10,10,10,10,10	0.965	0.943	0.974	0.943	0.713	0.974
	20,20,20,20,20	1.000	0.999	1.000	0.999	0.997	1.000
	30,30,30,30,30	1.000	1.000	1.000	1.000	1.000	1.000
	3,9,9,15,15	0.905	0.974	0.343	0.730	0.634	0.307
	5,15,15,20,20	0.981	0.994	0.711	0.956	0.948	0.695
	10,20,20,30,30	1.000	1.000	0.999	1.000	1.000	0.999
	15,15,9,9,3	0.774	0.735	0.825	0.540	0.052	0.813
	20,20,15,15,5	0.917	0.856	0.954	0.854	0.203	0.948
	30,30,20,20,10	0.991	0.979	0.995	0.988	0.856	0.995

0.25,0.25, 0.75, 1.25, 1.25	5,5,5,5,5	0.568	0.647	0.475	0.361	0.175	0.443
	10,10,10,10,10	0.824	0.819	0.818	0.751	0.503	0.812
	20,20,20,20,20	0.985	0.982	0.988	0.982	0.951	0.988
	30,30,30,30,30	0.999	0.999	0.999	0.999	0.997	0.999
	3,9,9,15,15	0.660	0.896	0.205	0.413	0.354	0.169
	5,15,15,20,20	0.820	0.937	0.421	0.704	0.704	0.393
	10,20,20,30,30	0.983	0.994	0.913	0.973	0.975	0.909
	15,15,9,9,3	0.623	0.594	0.684	0.361	0.033	0.658
	20,20,15,15,5	0.766	0.674	0.842	0.662	0.097	0.833
	30,30,20,20,10	0.933	0.870	0.958	0.908	0.593	0.954
0.5,0.5, 1.25, 2, 2	5,5,5,5,5	0.352	0.503	0.309	0.182	0.083	0.267
	10,10,10,10,10	0.495	0.535	0.511	0.397	0.199	0.495
	20,20,20,20,20	0.794	0.779	0.819	0.760	0.595	0.814
	30,30,30,30,30	0.937	0.931	0.946	0.930	0.864	0.945
	3,9,9,15,15	0.434	0.748	0.157	0.233	0.203	0.122
	5,15,15,20,20	0.525	0.759	0.256	0.395	0.380	0.233
	10,20,20,30,30	0.746	0.866	0.576	0.691	0.690	0.563
	15,15,9,9,3	0.392	0.420	0.473	0.200	0.027	0.427
	20,20,15,15,5	0.484	0.371	0.613	0.353	0.023	0.579
	30,30,20,20,10	0.654	0.514	0.751	0.592	0.180	0.739

As the LRT and WT have sizes higher than the nominal level, the values of these tests are not included in the power evaluation.

The ST and SCAT have remarkably the best powers among the others when the sample sizes are equal, especially for small equal sample sizes. They are followed by the LRT and WT, respectively. As the sample size increases, the ST and SCAT tests retain the highest power, while the other tests gradually converge to the powers exhibited by the ST and SCAT. When the variances are less heterogeneous, the power values of ST, LRCAT, SCAT, and WCAT are quite close to each other, especially at around n=20. However, when the difference between scale parameters increases, this convergence is not achieved even at n=30. This situation is consistently observed as the number of groups increases.

In situations where the sample sizes and scale parameters within each group are directly proportional, WCAT shows superior performance compared to the others. As the sample size increases, LRTCAT becomes quite close to WCAT in terms of power. Additionally, it is observed that LRCAT is slightly better than WCAT in cases of small sample sizes as the number of groups increases. Similarly to the case of equal sample sizes, when the variances are less heterogeneous, all tests converge faster as the sample size increases. However, this convergence is delayed when the difference between the variances increases.

In situations where the sample sizes and scale parameters within each group are inversely proportional, the ST and SCAT exhibit significantly better power values compared to the others. The LRCAT also performs better than the WCAT. In cases where the sample sizes increase and the variances are less heterogeneous, the LRCAT approaches the power values of the ST and SCAT.

However, the WCAT had difficulty in achieving this. When the difference between the variances increases, it is observed that both LRCAT and WCAT are more difficult to capture the power values of these tests. This situation showed similar results as the number of groups increased.

In general, based on our simulation results, the power values of all tests tend to be higher when the variances are less heterogeneous. Conversely, when heterogeneity increases, the power values of all tests tend to be lower. For instance, when we consider the case of k=4, θ=(0.1, 0.5, 0.5, 1), $\sigma^2 = (0.1, 0.5, 0.5, 1)$ and $n = (20,20,20,20)$, the powers of the ST, LRCAT, WCAT, SCAT are obtained as 0.962, 0.945, 0.896, 0.963, respectively. On the other hand, when k=4, θ=(0.1, 0.5, 0.5, 1), $\sigma^2 = (0.5, 1.25, 1.25, 2)$ and $n = (20,20,20,20)$, the powers of the ST, LRCAT, WCAT, SCAT are obtained as 0.451, 0.400, 0.293, 0.449, respectively. Additionally, the power values of all tests tend to rise as the difference between the values of θ increase.

4. Numerical example

In the study, the focus was on the survival of patients with advanced stomach, bronchial, colon, and ovarian cancer. Treatment with the ascorbate, which plays a role in the body's resistance to cancer, was used in patients. According to Cameron and Pauling [3], cancer patients have much lower levels of ascorbic acid, which is an observable biochemical property that suggests a significantly higher need and consumption of this substance to enhance these different host resistance characteristics. The results of a clinical trial in which patients with terminal cancer received supplementary ascorbate as part of normal treatment were presented. They found that the average survival time of the ascorbate subjects (more than 210 days) was more than 4.2 times that of the controls (50 days). We aim to explore whether survival times vary based on the affected organ. The data to be used for this analysis can be accessed from [7].

To evaluate the lognormality assumption for each group, the Shapiro-Wilk test, a normality test, was performed on the datasets after log transformation.

For each group, the p-values of this test are 0.965, 0.324, 0.961, and 0.189, respectively. These findings show that each group's dataset satisfies the assumption of lognormality. Table 10 shows the results of the summary statistics of the data for the four groups.

Table 10
The log- transformed survival times' summary statistics.

Groups	n_i	$\bar{x}_i$	s_i
Ovary	6	6.15	1.26
Stomach	13	4.97	1.25
Bronchus	17	4.95	0.95
Colon	17	5.75	0.99

Since the data are lognormally distributed, when the tests described in Section 2 are applied to test whether there is a difference between the mean

survival time for the 4 groups, the results in Table 12 are obtained. In addition, it should be noted that the sizes of the LRT and WT are higher than the nominal level, especially for small sample sizes. Therefore, we have not used the results of these tests, and the results of these tests are not reliable. Table 11 shows the *p*-values of all tests for the test of equality means of survival times for each cancer group. Note that using a Monte Carlo simulation with 10,000 'runs', these p-values were calculated.

Table 11
The result of all tests

Tests	ST	LRCAT	WCAT	SCAT
p-value	0.101	0.121	0.148	0.103

The Table 11 results show that the mean survival times of the four cancer groups are not statistically significantly different. Therefore, based on the nominal level of 5%, we can conclude that the patients' survival time does not vary according to the affected organ.

5. Conclusions

In this study, we investigated likelihood-based tests and suggested a new test, WT, for testing the means under log-normal populations. Since the likelihood-based tests are asymptotic tests, we also proposed bootstrap approaches for both WT and LRT. We also conducted Monte Carlo simulation experiments to compare the finite sample performance of these tests.

In the cases considered, it is observed that the sizes of the LRT and WT, except for the ST, are greater than the nominal level. The sizes of the bootstrap approaches of all tests show close to the nominal level. Regarding the power evaluation, different results are observed depending on the sample sizes and the relationship between sample size and scale parameters. In cases of equal sample sizes, the ST and SCAT exhibit notably superior performance, especially for small equal sample sizes. When the sample sizes and scale parameters in each group are directly proportional, WCAT outperforms other tests. Moreover, the LRCAT is slightly better than WCAT in the cases of small sample sizes as the number of groups increases. Conversely, when the sample sizes and scale parameters in each group are inversely proportional, the ST and SCAT exhibit significantly higher power values than the others. Finally, it is important to emphasize that researchers should consider using both WT and ST statistics, as well as LRT, in their studies.

References

[1] S. Boributh, R. Jiraratananon, and K. Li, "Analytical solutions for membrane wetting calculations based on log-normal and normal

distribution functions for CO2 absorption by a hollow fiber membrane contactor," *Journal of Membrane Science,* vol. 429, pp. 459-472 (2013).

[2] M. Cakmak, E. Gokpinar, F. Gokpinar, and M. Ebegil, "A revisit to testing the equality of means for several lognormal distributions," *Journal of Statistical Computation and Simulation*, vol. 90, no. 6, pp. 961-1000 (2020).

[3] E. Cameron and L. Pauling, "Supplemental ascorbate in the supportive treatment of cancer: reevaluation of prolongation of survival times in terminal human cancer," *Proceedings of the National Academy of Sciences*, vol. 75, no. 9, pp. 4538-4542 (1978).

[4] A. C. Davison and D. V. Hinkley, Bootstrap Methods and Their Application, no. 1. Cambridge University Press (1997).

[5] R. F. Engle and W. Wald, "Likelihood ratio and Lagrange multiplier tests in econometrics," in Handbook of Econometrics, Z. Griliches and M. D. Intriligator, Eds., vol. 2. Amsterdam: Elsevier, pp. 775–826 (1984).

[6] P. S. Gill, "Small-sample inference for the comparison of means of log-normal distributions," *Biometrics*, vol. 60, pp. 525–527 (2004).

[7] D. J. Hand, F. Daly, and K. McConway, A Handbook of Small Data Sets, vol. 1. Chapman and Hall/CRC Press, New York (1993).

[8] K. Krishnamoorthy and T. Mathew, "Inferences on the means of log-normal distributions using generalized p-values and generalized confidence intervals," *Journal of Statistical Planning and Inference,* vol. 115, no. 1, pp. 103-121 (2003).

[9] Y. T. N. Lee, "The lognormal distribution of growth rates of soft tissue metastases of breast cancer," *Journal of Surgical Oncology,* vol. 4, no. 2, pp. 81-88 (1972).

[10] X. Li, "A generalized p-value approach for comparing the means of several log-normal populations," *Statistics & Probability Letters*, vol. 79, no. 11, pp. 1404-1408 (2009).

[11] E. Limpert, W. A. Stahel, and M. Abbt, "Log-normal distributions across the sciences: keys and clues," *Bioscience*, vol. 51, pp. 341–352 (2001).

[12] S. H. Lin, "The higher order likelihood method for the common mean of several lognormal distributions," *Metrika*, vol. 76, pp. 381–392 (2013).

[13] M. Lubrano and A. A. J. Ndoye, "Income inequality decomposition using a finite mixture of log-normal distributions: A Bayesian approach," *Computational Statistics & Data Analysis*, vol. 100, pp. 830-846 (2016).

[14] N. Pal, W. K. Lim, and C. H. Ling, "A computational approach to statistical inferences," *Journal of Applied Probability and Statistics*, vol. 2, no. 1, pp. 13-35 (2007).

[15] J. Wu, G. Jiang, and A. Wong, "Likelihood analysis for the ratio of means of two independent log-normal distributions," *Biometrics*, vol. 58, pp. 463–469 (2002).

[16] J. C. W. Rayner, "The asymptotically optimal tests," *The Statistician*, vol. 3, pp. 337-346 (1997).

[17] L. D. Ribeiro-Reis, "The unit log-normal distribution: An alternative for beta and Kumaraswamy distributions," *Journal of Statistics and Management Systems,* vol. 27, no. 4, pp. 785–796 (2024).

[18] X. H. Zhou and W. Tu, "Comparison of several independent population means when their samples contain log-normal and possibly zero observations," Biometrics, vol. 55, pp. 645–651 (1999).

Received April, 2024