http://communications.science.ankara.edu.tr

Commun.Fac.Sci.Univ.Ank.Ser. A1 Math. Stat.
Volume 72, Number 3, Pages 686–700 (2023)
DOI:10.31801/cfsuasmas.1175872
ISSN 1303-5991 E-ISSN 2618-6470

COMMUNICATIONS SERIES A1

Research Article; Received: September 15, 2022; Accepted: March 20, 2023

COMPARISON OF THE SEVERAL TWO-PARAMETER EXPONENTIAL DISTRIBUTED GROUP MEANS IN THE PRESENCE OF OUTLIERS

Mustafa CAVUS[1] and Berna YAZICI[2]

[1,2]Eskisehir Technical University, Department of Statistics, Eskisehir, TÜRKİYE

Abstract. The two-parameter exponential distribution is often used to model the lifetime of a product. The comparison of the mean lifetimes of several products is a main concern in reliability applications. In this study, the performance of the methods to compare the mean lifetimes of several products based on generalized p-value, parametric bootstrap, and fiducial approach are compared in the presence of outliers. The results of Monte-Carlo simulations clearly indicate that there is no uniformly powerful test. The parametric bootstrap test is superior to the others except in the case of the lower number of groups and the presence of outliers. An illustrative example of testing the equality lifetimes of a component is given to perform the proposed tests. The considered tests are implemented in an R package `doex`.

1. Introduction

Testing equality of means of several normal populations under unequal variances is a very common Behrens-Fisher-type problem in social sciences, agriculture, biology, etc. The generalized p-value method is used to solve this problem [1]. The generalized F-test is proposed using the generalized p-value method, and its modifications for non-normality caused by outliers are improved by Cavus et al. [2], caused by skewness by Cavus et al. [3], and performed in a real data application by Cavus et al. [4]. Moreover, there are few parametric methods for testing the equality of means of skewed populations. Tian and Wu [5] proposed a generalized p-value approach for log-normal populations, Tian [6], Ma and Tian [7] improved procedures for inverse Gaussian and Niu et al. [8] proposed a generalized p-value procedure for Birbaum-Saunders distributions.

2020 *Mathematics Subject Classification.* 62J10, 62K99, 62F03.
Keywords. ANOVA, parametric bootstrap penalized power, exponential distribution, lifetime data.
[1] mustafacavus@eskisehir.edu.tr-Corresponding author; 0000-0002-6172-5449
[2] bbabloglu@eskisehir.edu.tr; 0000-0001-9843-7355.

The two-parameter exponential distribution is used in many real-life problems such as modeling extreme rainfalls, the lifetime of a component, the service time of an agent, and so on. Ghosh and Razmpour [9] indicated that two-parameter exponential distribution is used to model the guaranteed time with unknown and possibly unequal failure rates in reliability and life testing. There are some procedures improved for the two-parameter exponential distribution. Chen [10] proposed a range statistic for comparing location parameters of two-parameter exponential distributions. Singh [11] derived a likelihood ratio test for testing the equality of location parameters of two-parameter exponential distributions based on Type II censored samples under unknown scales. Kambo and Awad [12] proposed a test statistic based on doubly censored samples to test the equality of location parameters of k exponential distributions when the scale parameter is unknown. Hsieh [13] proposed an exact test for comparing location parameters simultaneously of several two-parameter exponential distributions under unequal scale parameters with unknown scale parameters. Vaughan and Tiku [14] extended the test developed by Tiku and Vaughan [15] for $k > 2$ populations for testing equality of location parameters of two-parameter exponential populations from censored samples. Ananda and Weerahandi [16] proposed a testing procedure based on generalized p-values for testing the difference between two exponential means. Wu [17] proposed a one-stage multiple comparison procedure for comparing $k - 1$ treatment exponential mean lifetimes with the control based on doubly censored samples under unequal scales. Malekzadeh and Jafari [18] proposed some procedures based on generalized p-values, parametric bootstrap, and fiducial approaches by using Cochran type test statistics for testing the means of several two-parameter exponential distributions under progressively Type II censoring. The two-parameter exponential distribution has scale and location parameters. In the testing equality of means of two-parameter exponential distributions, the scale parameter is a nuisance parameter when it is unknown or unequal. Therefore, the considered problem turns into a Behrens-Fisher-type problem. There is no study on the testing equality of two-parameter exponentially distributed population means for complete data in the presence of outliers.

The article discusses the testing equality means of k two-parameter exponentially distributed populations for complete data in the presence of outliers. In the next section, the procedures proposed by Malekzadeh and Jafari [18] are introduced. A Monte-Carlo simulation study is conducted for comparing the performances of these tests for complete data in the presence of outliers in Sec 3. To show the efficiency of the tests, illustrative examples are given in Sec 4. The results are discussed in the last section.

2. Methodology

In this section, methods proposed by Malekzadeh and Jafari [18] are introduced. The probability density function of the two-parameter exponential distribution is given in (1).

$$f(x;a,b) = \frac{1}{a}\exp\Big\{-\frac{x-b}{a}\Big\}, x > b, a > 0 \tag{1}$$

where a is the scale and b is the location parameter. We are interested in the problem of testing the equality of means of k exponentially distributed populations for complete data in (2).

$$\begin{aligned} &H_0 : \mu_1 = \mu_2 = ... = \mu_k \\ &H_A : \mu_i \neq \mu_j \text{ for some } i \text{ and } j \text{ where } i \neq j \end{aligned} \tag{2}$$

Rahman and Pearson [19] revisited the parameter estimations of two-parameter exponential distribution and conducted a simulation study to compare the performance of maximum likelihood, product spacing, and quantile estimation methods. The uniformly minimum variance unbiased estimators of the two-parameter exponential distribution parameters (Malekzadeh and Jafari, [18]):

$$\hat{a} = S/(n-1) \tag{3}$$

$$\hat{b} = X_{(1)} \tag{4}$$

where $X_{(1)} = min(X_1, X_2, ..., X_n)$ and $S = \sum_{j=1}^{n}[X_j - X_{(1)}]$. Viveros and Balakrishnan [20] gave the distributions of the following random variables.

$$W = \frac{2(n-1)S}{a} \sim \chi^2_{(2n-2)} \text{ and } Y = \frac{2n(X_{(1)}-b)}{a} \sim \chi^2_{(2)} \tag{5}$$

where W_i and Y_i are independent random variables. Cochran [21] type test statistics are used for Behrens-Fisher problems. Here, it is modified for testing the equality of two-parameter exponential distributed means under unequal scale parameters.

$$T_t = \sum_{i=1}^{k}\frac{n_i\hat{\mu}_i^2}{S_i^2} - \frac{\left(\sum_{i=1}^{k}\frac{n_i\hat{\mu}_i^2}{S_i^2}\right)^2}{\sum_{i=1}^{k}\frac{n_i}{S_i^2}} \tag{6}$$

where $\hat{\mu}$ is the mean estimate and S is the scale estimate of the i^{th} population. The uniformly minimum variance unbiased estimator of $\mu = a + b$ and it can be shown as in (7).

$$\hat{\mu}_i = X_{i(1)} + \frac{n_i - 1}{n_i} S_i = \frac{a_i}{2n_i}(W_i + Y_i) + b_i \sim N(\mu_i, a_i^2/n_i) \tag{7}$$

T_t is used for the rejection rule as a critical value of the Generalized p-value, Parametric Bootstrap, and Fiducial Approach test in the following subsections.

2.1. **Generalized p-value (GP) Based Test.** The generalized p-value method is used to derive the test statistics in the presence of nuisance parameters. Weerahandi [22] proposed the Generalized F-test for testing the equality of several populations' means under unequal variances instead of the Classical F-test. Also, many researchers used this method to derive test statistics for several distributions. In this method, firstly sufficient statistics of parameters of the related distribution are obtained. Using the sufficient statistics of the two-parameter exponential distribution, (i) R_i can be obtained independently from the nuisance parameter, and, (ii) since the observed λ_i values are independent of the nuisance parameter θ_i, generalized pivot value can be estimated.

$$R_i = X_{i(1)} + (n_i - 1)S_i(2n_i - Y_i/n_i W_i) \tag{8}$$

Expected values of $(X_{i(1)}, S_i)$ vector for R_i generalized pivot value and the variance can be obtained as follows:

$$\mu_{Ri} = X_{i(1)} + \frac{(n_i - 1)^2 S_i}{n_i^2 - 2n_i} \tag{9}$$

$$\sigma^2_{Ri} = \frac{(n_i - 1)^4 S_i^2}{n_i^2 (n_i - 2)^2} \left(\frac{1}{n_i - 3} \right) \tag{10}$$

Cochran test statistic can be obtained as in (11) using expected value of R_i generalized pivot and the variance of it.

$$T_{GP} = \sum_{i=1}^{k} \frac{(R_i - \mu_{Ri})^2}{\sigma^2_{Ri}} - \frac{\left(\sum_{i=1}^{k} \frac{R_i - \mu_{Ri}}{\sigma^2_{Ri}} \right)}{\sum_{i=1}^{k} \frac{1}{\sigma^2_{Ri}}} \tag{11}$$

The rejection rule is H_0 is in (2) rejected when $T_{GP} > T_t$. The p-value of the GP test is computed at least 10.000 Monte-Carlo runs as $p_{GP} = P(T_{GP} \geq T_t)$.

2.2. **Parametric Bootstrap (PB) Based Test.** Krishnamoorthy et al. [23] propose the parametric bootstrap method for testing the equality of normal population means under heteroscedasticity. Let $Y_i \sim \chi^2_{(2)}$ and $W_i \sim \chi^2_{(2n_i-2)}$. The PB test statistic is in (12) obtained for complete data from Malekzadeh and Jafari [18] using the Cochran statistic.

$$T_{PB} = \sum_{i=1}^{k} \frac{n_i \mu_{Bi}^2}{S_{Bi}^2} - \frac{\left(\sum_{i=1}^{k} \frac{n_i \mu_{Bi}^2}{S_{Bi}^2}\right)^2}{\sum_{i=1}^{k} \frac{n_i}{S_{Bi}^2}} \tag{12}$$

where $\mu_{Bi} = (S_i/2n_i)(W_i + Y_i)$ and $S_{Bi} = S_i W_i/(2n_i - 2)$. The rejection rule is H_0 is in (2) rejected when $T_{PB} > T_t$. The p-value of the PB test is computed at least 10.000 Monte-Carlo runs as $p_{PB} = P(T_{PB} \geq T_t)$.

2.3. **Fiducial Approach (FA) Based Test.** Li et al. [24] used the fiducial approach for testing the equality of several populations' means under unequal variances. Let $Y_i \sim \chi^2_{(2)}$ and $W_i \sim \chi^2_{(2n_i-2)}$, and S_i functions can be rewritten as random samples:

$$S_i = \frac{a_i W_i}{2(n_i - 1)}, \quad X_{i(1)} = \frac{a_i Y_i}{2n_i} + b_i \tag{13}$$

Parameter estimations are obtained as follows by using the observed values of $(X_{i(1)}, S_i)$

$$b_i = X_{i(1)} - \frac{(n_i - 1) S_i Y_i}{n_i W_i}, \quad a_i = \frac{2(n_i - 1) S_i}{W_i} \tag{14}$$

Using Cochran test statistic, T_{FA} can be written as in (15).

$$T_{FA} = \sum_{i=1}^{k} \frac{f_i n_i}{S_i n_i^2 W_i^2} - \frac{\left(\sum_{i=1}^{k} \frac{f_i}{S_i^2 n_i W_i}\right)^2}{\sum_{i=1}^{k} \frac{n_i}{S_i^2}} \tag{15}$$

where $f_i = (n_i - 1)(W_i Y_i - 2n_i W_i)$. The rejection rule is H_0 is in (2) rejected when $T_{PB} > T_t$. The p-value of the FA test is computed at least 10.000 Monte-Carlo runs as $p_{FA} = P(T_{FA} \geq T_t)$.

3. Monte-Carlo Simulation Study

In this section, we provide some of our comprehensive simulation study results. The GP, PB, and FA tests, as introduced in the previous subsections, are compared in terms of penalized power and Type I error probability when the nominal level of the test is taken as $\alpha_0 = 0.05$ under different sample sizes and scale parameters. The configuration of the outliers is determined similarly to the illustrative examples in the next section. The first and third groups consist of outlier one each which is five and three times higher than the group median, respectively in $k = 3$ groups design while the second, third, and fourth groups consist of an outlier one each which is one and a half times higher than the group median, respectively in $k = 4$ groups design.

It is known that Monte-Carlo simulation studies are used to compare the performance of the tests in terms of power and Type I error probability. However, any comparison of the powers is invalid when Type I error probabilities are different. Cavus et al. [25] proposed the penalized power approach in (16) to compare the power of the tests when Type I error probabilities are different.

$$\gamma_i = \frac{1-\beta_i}{\sqrt{1+|1-\frac{\alpha_i}{\alpha_0}|}} \tag{16}$$

where β_i is Type II error rate, α_i is Type I error of the test and α_0 is the nominal level. Penalized power adjusts the power function with the square root of the percentile deviation between Type I error probability and the nominal level. Thus, penalized power is used to compare the power of the tests in the simulation studies. The simulations are performed for balanced and unbalanced designs with `doex` package implemented by Cavus and Yazici [26] and Cavus and Yazici [27] in `R`, and the results are based on 10.000 Monte-Carlo runs. The results of the simulations are given in the following subsections.

3.1. **Type I Error Probability Results.** Table 1 shows the Type I error probabilities of the tests under scale parameters 2 and 5 for small, moderate, and large samples with and without outliers. The GP and FA test can not control Type I error probability in small samples for $\alpha_0 = 0.05$ while the PB test controls Type I error probability under unbalanced design $n_i = (5, 10, 15)$. The performance of the PB test to control the Type I error probability is not similar in the presence of outliers. It does not control the Type I error probability and shows a more conservative performance than the design without outliers. The FA and GP test generally has Type I errors close to each other and are more conservative than the PB test. In the presence of outliers, the performance of the GP and FA tests are affected negatively also and show more conservative performance. The PB test performs better than the others in moderate and large samples and controls the error. The performance of the GP and FA test is getting better in large and moderate samples. The performance of the test on controlling the Type I error probability is getting better when the number of groups (k) is increased. Even if the presence of outliers negatively affects the performance of all tests to control the Type I error probability, the increase in sample size eliminates this negative effect for GP and PB tests.

Table 1. Type I error probabilities for $\alpha_0 = 0.05$

				without outliers			with outliers		
k	n_i	a_i	b_i	GP	PB	FA	GP	PB	FA
3	10, 10, 10	2, 2, 2	1, 1, 1	0.0061	0.0087	0.0010	0.0130	0.0100	0.0005
		5, 5, 5		0.0065	0.0101	0.0014	0.0090	0.0095	0.0001
	8, 10, 12	2, 2, 2		0.0068	0.0128	0.0021	0.0140	0.0030	0.0008
		5, 5, 5		0.0083	0.0139	0.0025	0.0130	0.0030	0.0002
	5, 10, 15	2, 2, 2		0.0144	0.0436	0.0072	0.0120	0.0003	0.0009
		5, 5, 5		0.0129	0.0439	0.0079	0.0070	0.0003	0.0008
	30, 30, 30	2, 2, 2		0.0269	0.0407	0.0201	0.0410	0.0580	0.0310
		5, 5, 5		0.0298	0.0462	0.0223	0.0330	0.0540	0.0280
	24, 30, 36	2, 2, 2		0.0309	0.0453	0.0236	0.0440	0.0530	0.0260
		5, 5, 5		0.0298	0.0450	0.0229	0.0350	0.0500	0.0210
	15, 30, 45	2, 2, 2		0.0332	0.0487	0.0242	0.0380	0.0390	0.0220
		5, 5, 5		0.0325	0.0528	0.0248	0.0350	0.0320	0.0150
	50, 50, 50	2, 2, 2		0.0365	0.0473	0.0310	0.0480	0.0650	0.0440
		5, 5, 5		0.0348	0.0455	0.0302	0.0432	0.0604	0.0360
	40, 50, 60	2, 2, 2		0.0359	0.0493	0.0319	0.0490	0.0570	0.0360
		5, 5, 5		0.0368	0.0498	0.0309	0.0470	0.0540	0.0320
	25, 50, 75	2, 2, 2		0.0392	0.0494	0.0308	0.0460	0.0465	0.0320
		5, 5, 5		0.0407	0.0512	0.0339	0.0430	0.0410	0.0283
4	10, 10, 10, 10	2, 2, 2, 2	1, 1, 1, 1	0.0054	0.0094	0.0007	0.0070	0.0100	0.0010
		5, 5, 5, 5		0.0052	0.0090	0.0006	0.0060	0.0100	0.0010
	7, 9, 11, 13	2, 2, 2, 2		0.0086	0.0165	0.0029	0.0070	0.0170	0.0030
		5, 5, 5, 5		0.0083	0.0162	0.0027	0.0070	0.0170	0.0030
	5, 8, 12, 15	2, 2, 2, 2		0.0120	0.0337	0.0059	0.0050	0.0360	0.0080
		5, 5, 5, 5		0.0121	0.0332	0.0056	0.0050	0.0360	0.0080
	30, 30, 30, 30	2, 2, 2, 2		0.0256	0.0404	0.0183	0.0270	0.0450	0.0220
		5, 5, 5, 5		0.0252	0.0399	0.0180	0.0280	0.0460	0.0210
	21, 27, 33, 39	2, 2, 2, 2		0.0320	0.0486	0.0218	0.0250	0.0430	0.0220
		5, 5, 5, 5		0.0310	0.0481	0.0213	0.0280	0.0440	0.0220
	15, 24, 36, 45	2, 2, 2, 2		0.0319	0.0489	0.0245	0.0260	0.0510	0.0300
		5, 5, 5, 5		0.0314	0.0482	0.0241	0.0260	0.0520	0.0300
	50, 50, 50, 50	2, 2, 2, 2		0.0317	0.0444	0.0273	0.0360	0.0490	0.0300
		5, 5, 5, 5		0.0314	0.0442	0.0271	0.0360	0.0500	0.0320
	35, 45, 55, 65	2, 2, 2, 2		0.0312	0.0443	0.0264	0.0330	0.0500	0.0320
		5, 5, 5, 5		0.0310	0.0442	0.0261	0.0340	0.0520	0.0310
	25, 40, 60 ,75	2, 2, 2, 2		0.0360	0.0492	0.0299	0.0290	0.0460	0.0320
		5, 5, 5, 5		0.0350	0.0488	0.0295	0.0290	0.0450	0.0320

3.2. **Penalized Power Results.** Table 2 shows the results of penalized powers of the tests in the case of $k = 3$ for several effect sizes and sample sizes. Recall from Table 1 that the GP and FA tests are very conservative in terms of Type I error probability, while the PB test successfully controls the Type I error probability.

The penalized power results show that the PB test is more powerful than the GP and FA test in most of the scenarios except the case of unbalanced small sample size designs. In higher effect sizes for large samples, penalized power of the tests are higher than 0.85. Also, their performances are better in unbalanced designs than in balanced designs. The performance of the GP and PB tests is affected negatively when the scale parameter is increased while the performance of the FA test is positively affected without outliers. It is seen that the power of the tests decreases in the case of $\theta_i = 5$. For example, the power of the PB test is 0.99 in the case of $\theta_i = 3$ and 0.96 in $\theta_i = 5$, it is the biggest difference between the tests. It is concluded that the effect of the higher scale parameter on the PB test is higher than the others. However, the penalized power of the PB test is the highest in most of the scenarios followed by the GP test and the FA test. When the power of the tests is evaluated according to whether there is an outlier or not, it is seen that the GP and FA tests are higher in the case of outliers than in the case of no outliers, and the contrary, the power of the PB test is lower. The result is that PB is the uniformly most powerful test in the non-presence of an outlier, and GP is the uniformly most powerful test in the case of an outlier.

Table 3 shows the results of penalized powers for $k = 4$. Unlike the results in Table 2, the most powerful test is the PB, the second is GP and the last one is the FA test in the presence and non-presence of outliers. The increase in the number of groups affects the penalized power of the tests negatively in small samples in most of the scenarios. Only the performance of the PB test is better than the case of $k = 3$ in large samples and it is obtained that the least affected test is the PB test.

When the results given in Tables 2 and 3 are examined, the effect of the design configurations such as the presence of outliers and the number of groups on the performance of the tests differs. Therefore, when using tests, the reliability of their results should be carefully examined.

4. Illustrative Examples

In this section, the GP, PB, and FA tests are applied to two real data examples to compare their results in hypothesis testing.

Example 1. Data consists of the lifetimes of a component are different brands in a refrigerator which is collected from a local factory in Turkey and it is available in `doex` package in `R` as component data. It is known that the lifetime data generally follows the exponential distribution. However, to make sure of this, the Cramer-von Mises (CvM) goodness-of-fit test is used to test whether the data follows the two-parameter exponential distribution. As a result of the CvM test, the p-value 0.6786 shows there is not enough evidence to reject the null hypothesis indicating that the data follows a two-parameter exponential distribution at the 0.05 significance level. The sample size of the data is $n_1 = 15, n_2 = 49, n_3 = 54, n_4 = 12$. The estimates of the location parameters are $\hat{b}_1 = 8.38, \hat{b}_2 = 8.40, \hat{b}_3 = 8.41, \hat{b}_4 = 8.62$ and the

TABLE 2. Penalized power results for $k = 3$

			without outliers			with outliers		
n_i	a_i	b_i	GP	PB	FA	GP	PB	FA
10, 10, 10	2, 2, 3	1, 1, 1	0.0101	0.0143	0.0021	0.0280	0.0193	0.0028
	2, 2, 4		0.0254	0.0334	0.0055	0.0515	0.0432	0.0070
	2, 2, 5		0.0429	0.0581	0.0097	0.0864	0.0648	0.0155
8, 10, 12	2, 2, 3		0.0245	0.0342	0.0039	0.0434	0.0150	0.0281
	2, 2, 4		0.0622	0.0779	0.0096	0.1006	0.0358	0.0091
	2, 2, 5		0.1181	0.1386	0.0177	0.1890	0.0746	0.0141
5, 10, 15	2, 2, 3		0.0559	0.1119	0.0147	0.0889	0.0028	0.0001
	2, 2, 4		0.1590	0.2180	0.0265	0.2231	0.0042	0.0002
	2, 2, 5		0.3010	0.3475	0.0413	0.3572	0.0106	0.0004
30, 30, 30	2, 2, 3		0.1610	0.2420	0.1253	0.2062	0.2748	0.1489
	2, 2, 4		0.4997	0.6363	0.4303	0.5790	0.6685	0.4860
	2, 2, 5		0.7297	0.8488	0.6697	0.8220	0.8718	0.7320
24, 30, 36	2, 2, 3		0.2444	0.3225	0.1755	0.3061	0.3418	0.1824
	2, 2, 4		0.6301	0.7455	0.5232	0.7238	0.7653	0.5392
	2, 2, 5		0.8083	0.9180	0.7411	0.8976	0.9285	0.7587
15, 30, 45	2, 2, 3		0.3307	0.4043	0.2160	0.3681	0.3023	0.1905
	2, 2, 4		0.7178	0.8277	0.5868	0.7696	0.7360	0.5572
	2, 2, 5		0.8447	0.9643	0.7630	0.8791	0.8800	0.7429
50, 50, 50	2, 2, 3		0.3808	0.4686	0.3422	0.4324	0.4323	0.3968
	2, 2, 4		0.8119	0.9107	0.7675	0.8835	0.8068	0.8400
	2, 2, 5		0.8812	0.9700	0.8447	0.9776	0.8761	0.9420
40, 50, 60	2, 2, 3		0.4722	0.5521	0.4015	0.5069	0.4842	0.3836
	2, 2, 4		0.8448	0.9522	0.8029	0.9356	0.8860	0.8158
	2, 2, 5		0.8821	0.9921	0.8549	0.9881	0.9347	0.8812
25, 50, 75	2, 2, 3		0.5582	0.6071	0.4440	0.5831	0.5263	0.3910
	2, 2, 4		0.8840	0.9662	0.8103	0.9343	0.9276	0.8077
	2, 2, 5		0.9063	0.9937	0.8489	0.9593	0.9593	0.8532
10, 10, 10	5, 5, 6	1, 1, 1	0.0215	0.0234	0.0164	0.0133	0.0111	0.0007
	5, 5, 8		0.0298	0.0321	0.0266	0.0266	0.0200	0.0042
	5, 5, 10		0.0419	0.0436	0.0369	0.0385	0.0422	0.0070
8, 10, 12	5, 5, 6		0.0356	0.0367	0.0349	0.0121	0.0071	0.0003
	5, 5, 8		0.0651	0.0642	0.0622	0.0447	0.0172	0.0028
	5, 5, 10		0.1046	0.1051	0.1023	0.0932	0.0394	0.0084
5, 10, 15	5, 5, 6		0.0221	0.0631	0.0094	0.0256	0.0002	0.0001
	5, 5, 8		0.0726	0.1325	0.0171	0.0997	0.0007	0.0003
	5, 5, 10		0.1577	0.2186	0.0266	0.1994	0.0063	0.0005
30, 30, 30	5, 5, 6		0.1509	0.1628	0.1329	0.0587	0.0991	0.0458
	5, 5, 8		0.4060	0.4592	0.3732	0.2531	0.3791	0.1975
	5, 5, 10		0.5755	0.6553	0.5404	0.5425	0.6908	0.4783
24, 30, 36	5, 5, 6		0.0758	0.0868	0.0742	0.0789	0.1090	0.0477
	5, 5, 8		0.2967	0.3371	0.2831	0.3552	0.4340	0.2426
	5, 5, 10		0.5172	0.5825	0.4921	0.6718	0.7910	0.5234
15, 30, 45	5, 5, 6		0.0886	0.1306	0.0588	0.1157	0.0866	0.0467
	5, 5, 8		0.4225	0.4979	0.2934	0.4481	0.3798	0.2316
	5, 5, 10		0.7141	0.8159	0.5891	0.7481	0.7022	0.5307
50, 50, 50	5, 5, 6		0.2984	0.3276	0.2862	0.1058	0.1314	0.0848
	5, 5, 8		0.6130	0.6734	0.5938	0.5366	0.5741	0.4347
	5, 5, 10		0.6634	0.7277	0.6399	0.8419	0.8389	0.7866
40, 50, 60	5, 5, 6		0.1163	0.1323	0.1123	0.1233	0.1299	0.0771
	5, 5, 8		0.4950	0.5549	0.4724	0.6313	0.6350	0.4956
	5, 5, 10		0.6630	0.7370	0.6277	0.9188	0.9160	0.7914
25, 50, 75	5, 5, 6		0.1498	0.1752	0.1097	0.1760	0.1445	0.0958
	5, 5, 8		0.6815	0.7300	0.5755	0.7024	0.6379	0.5091
	5, 5, 10		0.8951	0.9605	0.8291	0.9122	0.8920	0.7883

Table 3. Penalized power results for $k = 4$

			without outliers			with outliers		
n_i	a_i	b_i	GP	PB	FA	GP	PB	FA
10, 10, 10, 10	2, 2, 2, 3	1, 1, 1, 1	0.0086	0.0130	0.0016	0.0102	0.0134	0.0021
	2, 2, 2, 4		0.0149	0.0262	0.0028	0.0146	0.0283	0.0049
	2, 2, 2, 5		0.0227	0.0406	0.0050	0.0175	0.0402	0.0063
7, 9, 11, 13	2, 2, 2, 3		0.0196	0.0319	0.0037	0.0117	0.3182	0.0050
	2, 2, 2, 4		0.0554	0.0687	0.0075	0.0263	0.0651	0.0100
	2, 2, 2, 5		0.1074	0.1142	0.0112	0.0703	0.1086	0.0122
5, 8, 12, 15	2, 2, 2, 3		0.0417	0.0651	0.0090	0.0224	0.0707	0.0103
	2, 2, 2, 4		0.1214	0.1242	0.0133	0.0783	0.1149	0.0154
	2, 2, 2, 5		0.2425	0.1924	0.0177	0.1784	0.1829	0.0243
30, 30, 30, 30	2, 2, 2, 3		0.1254	0.2023	0.0893	0.1191	0.2011	0.0880
	2, 2, 2, 4		0.4084	0.5593	0.3302	0.4063	0.5987	0.3226
	2, 2, 2, 5		0.6623	0.8024	0.5852	0.6728	0.8552	0.6108
21, 27, 33, 39	2, 2, 2, 3		0.2598	0.3147	0.1527	0.2490	0.3175	0.1457
	2, 2, 2, 4		0.6636	0.7661	0.4834	0.6385	0.7258	0.4931
	2, 2, 2, 5		0.8200	0.9441	0.7182	0.7928	0.9075	0.7229
15, 24, 36, 45	2, 2, 2, 3		0.3263	0.3573	0.1700	0.3041	0.3584	0.1656
	2, 2, 2, 4		0.7321	0.8051	0.5376	0.7052	0.8178	0.5628
	2, 2, 2, 5		0.8416	0.9591	0.7532	0.8104	0.9693	0.7885
50, 50, 50, 50	2, 2, 2, 3		0.3099	0.4026	0.2720	0.3394	0.4455	0.2974
	2, 2, 2, 4		0.7457	0.8605	0.7083	0.7919	0.9158	0.7428
	2, 2, 2, 5		0.8467	0.9413	0.8191	0.8785	0.9861	0.8383
35, 45, 55, 65	2, 2, 2, 3		0.4899	0.5281	0.3750	0.4941	0.5520	0.3918
	2, 2, 2, 4		0.8284	0.9155	0.7753	0.8396	0.9650	0.8094
	2, 2, 2, 5		0.8508	0.9459	0.8213	0.8630	0.9990	0.8566
25, 40, 60 75	2, 2, 2, 3		0.5669	0.5865	0.4151	0.5412	0.5725	0.4347
	2, 2, 2, 4		0.8643	0.9664	0.8084	0.8299	0.9439	0.8309
	2, 2, 2, 5		0.8770	0.9878	0.8415	0.8391	0.9622	0.8574
10, 10, 10, 10	5, 5, 5, 6	1, 1, 1, 1	0.0058	0.0077	0.0008	0.0051	0.0111	0.0007
	5, 5, 5, 8		0.0102	0.0154	0.0017	0.0123	0.0149	0.0028
	5, 5, 5, 10		0.0149	0.0262	0.0028	0.0145	0.0290	0.0049
7, 9, 11, 13	5, 5, 5, 6		0.0099	0.0196	0.0026	0.0080	0.0209	0.0028
	5, 5, 5, 8		0.0254	0.0381	0.0045	0.0139	0.0388	0.0057
	5, 5, 5, 10		0.0554	0.0686	0.0075	0.0293	0.0667	0.0100
5, 8, 12, 15	5, 5, 5, 6		0.0160	0.0406	0.0060	0.0094	0.0477	0.0081
	5, 5, 5, 8		0.0539	0.0758	0.0098	0.0340	0.0786	0.0117
	5, 5, 5, 10		0.1207	0.1244	0.0133	0.0834	0.1131	0.0154
30, 30, 30, 30	5, 5, 5, 6		0.0356	0.0649	0.0255	0.0408	0.0692	0.0326
	5, 5, 5, 8		0.1697	0.2679	0.1265	0.1608	0.2780	0.1209
	5, 5, 5, 10		0.4089	0.5607	0.3302	0.4100	0.6100	0.3222
21, 27, 33, 39	5, 5, 5, 6		0.0703	0.1044	0.0413	0.0691	0.0992	0.0440
	5, 5, 5, 8		0.3486	0.4173	0.2074	0.3508	0.4204	0.2113
	5, 5, 5, 10		0.6680	0.7676	0.4840	0.6558	0.7370	0.4963
15, 24, 36, 45	5, 5, 5, 6		0.0828	0.1161	0.0473	0.0723	0.1108	0.0532
	5, 5, 5, 8		0.4269	0.4626	0.2380	0.4035	0.4696	0.2484
	5, 5, 5, 10		0.7331	0.8067	0.5380	0.7069	0.8109	0.5654
50, 50, 50, 50	5, 5, 5, 6		0.0707	0.1042	0.0590	0.0830	0.1240	0.0728
	5, 5, 5, 8		0.4171	0.5298	0.3764	0.4498	0.5810	0.3999
	5, 5, 5, 10		0.7452	0.8605	0.7103	0.7910	0.9260	0.7554

estimates of the scale parameters are $\hat{a}_1 = 1.47, \hat{a}_2 = 1.60, \hat{a}_3 = 1.82, \hat{a}_4 = 1.80$, respectively. It is clearly seen that the scale parameters are different. The lifetimes of the brands are given in Figure 1. The boxplots show that the groups referenced as Brands 2-4 consist of outliers. These outliers are higher than one and a half times higher than the medians. Testing the mean lifetimes of the components under scale parameters, GP, PB, and FA tests are performed by using the `doex`.
The p-value of the GP, PB, and FA tests are $0.6807, 0.7471$, and 0.7545, respectively. Thus, there is no evidence to reject the null hypothesis at $\alpha_0 = 0.05$ and concluded that the mean lifetimes of the components produced by different brands are not different. It is seen that the PB test can control the Type I error probability very close to the nominal level, in the unbalanced moderate, low-scale parameter and outlier design in Table 1. Therefore, it can be said that the results obtained in this example are reliable.

Example 2. In this example, the equality of mean agricultural income of the geographical regions in Turkey is considered. Agricultural incomes of the Central Anatolia (CA), Eastern Anatolia (EA), and Southeastern Anatolia (SA) regions in 2017 are considered and the data is obtained from the Turkish Statistical Institute Database. The Cramer-von Mises (CvM) goodness-of-fit test is used to test whether the data follows the two-parameter exponential distribution. As a result of the CvM test, the p-value 0.4005 shows there is not enough evidence to reject the null hypothesis indicating that the data follows a two-parameter exponential distribution at the 0.05 significance level. The number of city in the geographical regions are $n_{CA} = 13, n_{EA} = 14$, and $n_{SA} = 9$. The estimates of the location parameters are $\hat{b}_{CA} = 0.7503, \hat{b}_{EA} = 0.3649, \hat{b}_{SA} = 0.5811$, and the estimates of the scale parameters are $\hat{a}_{CA} = 2.2122, \hat{a}_{EA} = 1.0558, \hat{a}_{SA} = 1.7988$, respectively. The agricultural income of the geographical regions in Turkey is given in Figure 2. The boxplots show that the groups referenced as CA and SA consist of outliers. The outlier in the geographical region of CA is five times higher than the median while the outlier in the geographical region of SA is three times higher than its median. Testing the mean income of the geographical regions under unequal scale parameters, GP, PB, and FA tests are performed.
The p-value of the GP, PB, and FA tests are $0.0816, 0.0881$, and 0.1489, respectively. Thus, there is enough evidence to reject the null hypothesis at $\alpha_0 = 0.10$ and concluded that the mean incomes of the geographical regions are not different according to the results of the GP and PB test. In Table 2, the GP and PB tests are more powerful than the FA test, that's why it can be said that the results of these two tests are more reliable than the FA test in the presence of outliers.

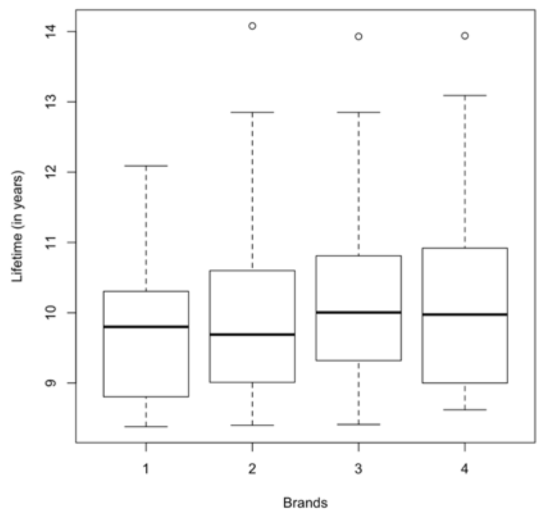

FIGURE 1. Lifetime of the components in years

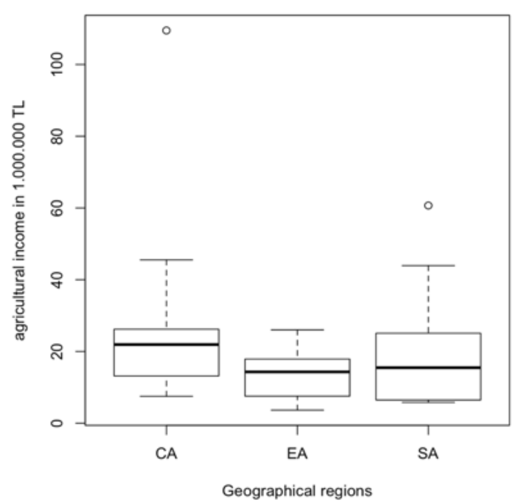

FIGURE 2. Agricultural income of the geographical regions in Turkey

5. Results and Conclusions

The generalized p-value, parametric bootstrap, and fiducial approach-based test proposed by Malekzadeh and Jafari [18] can be used for complete data. The performance of the tests was compared in terms of Type I error probability and penalized power for complete data and the most powerful test is determined. The results are obtained in balanced and unbalanced designs for small, moderate, and large samples in the presence of outliers. The simulation results clearly show that the PB test is superior to the others to control the Type I error probability and penalized power in most of the cases. Only in the presence of outliers, the GP test is more powerful than the PB test in $k = 3$ group designs. There are also some interesting results obtained such as the negative effect of the balanced designs and higher scale parameters on the performance of the tests. Moreover, illustrative examples are given to perform the tests on a real data example. It is concluded that the lifetimes of the components are not statistically significant. In this study, the PB test is obtained as a powerful test for testing the equality of exponentially distributed populations' means under unequal scale parameters and it can be safely used in reliability analysis, modeling extreme events, sequential analysis, and income inequality.

Author Contribution Statements The authors contributed equally. All authors read and approved the final copy of the paper.

Declaration of Competing Interests The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Tsui, K., Weerahandi, S., Generalized p-values in significance testing of hypotheses in the presence of nuisance parameters, *Journal of the American Statistical Association*, 84 (1989), 602–607. https://doi.org/10.2307/2289949

[2] Cavus, M., Yazici, B., Sezer, A., Modified tests for comparison of group means under heteroskedasticity and non-normality caused by outlier(s), *Hacettepe Journal of Mathematics and Statistics*, 46 (2017), 602–607. https://doi.org/10.15672/HJMS.2017.417

[3] Cavus, M., Yazici, B., Sezer, A., A revised generalized F-test for testing the equality of group means under non-normality caused by skewness, *Communications Faculty of Sciences University of Ankara Series A1 Mathematics and Statistics*, 70 (2021), 1036–1054. https://doi.org/10.31801/cfsuasmas.800452

[4] Cavus, M., Yazici, B., Sezer, A., Analyzing regional export data by the modified generalized F-test, *International Journal of Economic and Administrative Studies*, (2018), 541–552. https://doi.org/10.18092/ulikidince.348070

[5] Tian, L., Wu, J., Inferences on the common mean of several log-normal populations: the generalized variable approach, *Biometrical Journal*, 49 (2007), 944–951. https://doi.org/10.1002/bimj.200710391

[6] Tian, L., Testing equality of inverse gaussian means under heterogeneity based on generalized test variable, *Computational Statistics and Data Analysis*, 51 (2006), 1156–1162. https://doi.org/10.1016/j.csda.2005.11.012

[7] Ma, C., Tian, L., A parametric bootstrap approach for testing equality of inverse gaussian means under heterogeneity, *Communications in Statistics-Simulation and Computation*, 38 (2009), 1156–1162. https://doi.org/10.1080/03610910902833470

[8] Niu, C., Guo, X., Xu, W., Zhu, L., Comparison of several Birbaum-Saunders distributions, *Journal of Statistical Computation and Simulation*, 84 (2014), 2721–2733. https://doi.org/10.1080/00949655.2014.881814

[9] Ghosh, M., Razmpour, A., Estimation of the common location parameter of several exponentials, *Sankhya: The Indian Journal of Statistics*, 46 (1984), 383–394. https://www.jstor.org/stable/25050498

[10] Chen, H., A new range statistic for comparisons of several exponential location parameters, *Biometrika*, 69 (1982), 257–260. https://doi.org/10.2307/2335881

[11] Singh, N., The likelihood ratio test for the equality of location parameters of exponential populations based on Type II censored samples, *Technometrics*, 25 (1983), 193–195. https://doi.org/10.1080/00401706.1983.10487852

[12] Kambo, N., Awad, A., Testing equality of location parameters of k exponential distributions, *Communications in Statistics - Theory and Methods*, 14 (1985), 567–583.

[13] Hsieh, H., An exact test for comparing location parameters of k exponential distributions with unequal scaled based on Type II censored data, *Technometrics*, 28 (1986), 157–164. https://doi.org/10.2307/1270452

[14] Vaughan, D., Tiku, M., Testing the equality of location parameters of exponential populations from censored samples, *Communication in Statistics-Theory and Methods*, 22 (1993), 2567–2581. https://doi.org/10.1080/03610928308831169

[15] Tiku, M. Vaughan, D., Testing equality of location parameters of two exponential distributions from censored samples, *Communication in Statistics-Theory and Methods*, 20 (1991), 929–944. https://doi.org/10.1080/03610929108830540

[16] Ananda, M., Weerahandi, S., Testing the difference of two exponential means using generalized p-values, *Communications in Statistics - Simulation and Computation*, 25 (1996), 521–532. https://doi.org/10.1080/03610919608813327

[17] Wu, S., One stage multiple comparisons with the control for exponential mean lifetimes based on doubly censored samples under heteroscedasticity, *Communications in Statistics - Simulation and Computation*, 50 (2021), 1473-1483. https://doi.org/10.1080/03610918.2019.1584302

[18] Malekzadeh, A., Jafari, A., Inference on the equality means of several two-parameter exponential distributions under progressively Type II censoring, *Communications in Statistics - Simulation and Computation*, 49 (2020), 3196-3211. https://doi.org/10.1080/03610918.2018.1538452

[19] Rahman, M., Pearson, L. M., Estimation in two-parameter exponential distributions, *Journal of Statistical Computation and Simulation*, 70 (2001), 371–386. https://doi.org/10.1080/00949650108812128

[20] Viveros, R., Balakrishnan, N., Interval estimation of parameters of life from progressively censored data, *Technometrics*, 36 (1994), 84–91. https://doi.org/10.2307/1269201

[21] Cochran, W. G., Problem arising in the analysis of a series of similar experiments, *Journal of the Royal Statistical Society*, 4 (1937), 102–118. https://www.jstor.org/stable/2984123

[22] Weerahandi, S., ANOVA under unequal error variances, *Biometrics*, 51 (1995), 102–118. https://doi.org/10.2307/2532947

[23] Krishnamoorthy, K., Lu, F., Mathew, T., A parametric bootstrap approach for ANOVA with unequal variances: fixed and random models, *Computational Statistics and Data Analysis*, 51 (2007), 5731–5742. https://doi.org/10.1016/j.csda.2006.09.039

[24] Li, X., Wang, J., Liang, H., Comparison of several means: A fiducial based approach, *Computational Statistics and Data Analysis*, 55 (2011), 1993–2002. https://doi.org/10.1016/j.csda.2010.12.009

[25] Cavus, M., Yazici, B., Sezer, A., Penalized power approach to compare the power of the tests when Type I error probabilities are different, *Communications in Statistics - Simulation and Computation*, 50 (2021), 1912-1926. https://doi.org/10.1080/03610918.2019.1588310

[26] Cavus, M., Yazici, B., Testing the equality of normal distributed and independent groups' means under unequal variances by doex package, *The R Journal*, 12 (2021), 134-154. https://doi.org/10.32614/RJ-2021-008

[27] Cavus, M., Yazici, B., doex: The one-way heteroscedastic ANOVA tests, *R package*, v.1.2.