Hacettepe Journal of Mathematics and Statistics
Volume 46 (2) (2017), 339 – 360

# Modified generalized p-value and confidence interval by Fisher's fiducial approach

Evren Ozkip*† , Berna Yazici‡ and Ahmet Sezer§

**Abstract**

In this study, we develop two simple generalized confidence intervals for the difference between means of two normal populations with heteroscedastic variances which is usually referred to as the Behrens-Fisher problem. The developed confidence intervals are compared with the generalized confidence interval in the literature. We also propose modified fiducial based approach using Fisher's fiducial inference for comparing the mean of two lognormal distributions and compare them with the other tests in the literature. A Monte Carlo simulation study is conducted to evaluate performances of the proposed methods under different scenarios. The simulation results indicate that the developed confidences intervals for the Behrens-Fisher problem have shorter interval lengths and they give better coverage accuracy in some cases. The modified fiducial based approach is the best to provide satisfactory results in respect to its type error and power in all sample sizes. The modified test is applicable to small samples and is easy to compute and implement. The methods are also applied to two real-life examples.

**Keywords:** Modified fiducial based test, Generalized approach, Parametric bootstrap approach, Lognormal distribution, The Behrens-Fisher problem.

*2000 AMS Classification:* 60E05

*Received* : 23.06.2016 *Accepted* : 13.08.2016 *Doi* : 10.15672/HJMS.2017.412

*Ministry of National Education, Ankara, Turkey.,
Email: evrenozkip@gmail.com

†Corresponding Author.

‡Department of Statistics, Science Faculty, Anadolu University, Eskisehir, Turkey.,
Email: bbaloglu@anadolu.edu.tr

§Department of Statistics, Science Faculty, Anadolu University, Eskisehir, Turkey.,
Email: a.sezer@anadolu.edu.tr

## 1. Introduction

In practice, it is often of interests to compare the means of two populations, which are approximately normally distributed but with possibly different variances. When the variances are unknown, it is usually referred to as the Behrens-Fisher (BF) problem. The BF problem has been studied since the early 1930's. One reason for its popularity is that there is no exact solution satisfying the classical criteria for good tests. There have been quite few solutions proposed for the BF problem; see [1, 2, 3, 4, 5, 6, 7]. Kim and Cohen [8] presented a review of fundamental concepts and applications used to address the BF problem under fiducial, Bayesian, and frequentist approaches. Singh et al. [9] proposed a new test using Jackknife methodology. Dong [10] considered the empirical likelihood approach for the BF problem.

The idea is an extension of the solution for testing the BF problem, which was proposed by Tsui and Weerahandi [11] using the concept of the generalized p-values. Weerahandi [12] gave the concept of generalized pivotal quantities for constructing confidence intervals of scalar parameters, which are called generalized confidence intervals (GCI). Generalized confidence intervals were achieved to obtain exact confidence intervals in statistical problems involving nuisance parameters. Generalized procedures have been successfully applied to several problems of practical importance; see [13, 14, 15, 16, 17].

In recent years, Chang and Pal [18] revisited BF problem and apply a newly developed Computational approach test (CAT) that does not require explicit knowledge of the sampling distribution of the test statistic. Zheng et al. [19] proposed a two-stage method for BF problem. Ozkip et al. [20] presented a simulation study on some tests for the BF problem. Sezer et al. [21] compared three approximate confidence intervals and a generalized confidence interval for the BF problem for the two population case. They also showed how to obtain simultaneous confidence intervals for the 3 population case (ANOVA) by the Bonferroni correction factor. Ye et al. [22] considered the hypothesis testing and interval estimation for the reliability parameter in balanced and unbalanced one-way random models. They developed the tests and confidence intervals for the reliability parameter using the concepts of generalized $p$-value and generalized confidence interval. Gunasekera and Ananda [23] considered the development of inferential techniques based on the generalized variable method for the location parameter of the general half-normal distribution. Zhao and Xu [24] applied the generalized inference to the calibration problem, and took the generalized p-value as the test statistic to develop a new p-value for one-sided hypothesis testing, which they referred to as the one-sided posterior predictive $p$-value.

Faced with positive right-skewed data, the lognormal distribution is frequently used. It is difficult to construct exact tests and confidence intervals for comparing the mean of two lognormal distributions since the nuisance parameter is present. A major motivation for the present work is to provide inference procedure for the small samples when the interest is to compare means of two lognormal distributions. The lognormal distribution is frequently used for analyzing biological, medical and industrial data (Krishnamoorthy et al. [25] and Shen et al. [26]). Let X be a lognormal distributed random variable and $\mu$ and $\sigma^2$ denote the mean and variance of $lnX$, respectively, so that $Y = lnX \sim N(\mu, \sigma^2)$. The mean of $Y$ random variable depends on $\mu$ and $\sigma^2$ as expressed in equation 1.1.

$$(1.1) \qquad E(Y) = E(exp(X)) = exp(\eta) \quad \text{where} \quad \eta = \mu + \frac{\sigma^2}{2}$$

The hypothesis tests and confidence intervals about $X$ random variable depends on calculations corresponding to value of the nuisance parameter $\eta$. It is difficult to obtain exact tests and confidence intervals about $Y$ random variable in the presence of the nuisance parameter $\eta$. The problem of obtaining confidence intervals and tests concerning $\eta$ has been addressed by Land [27, 28, 29] and Angus [30, 31]. Zhou et al. [32] proposed a likelihood-based approach and a bootstrap-based approach. Wu et al. [33] proposed the signed log-likelihood ratio statistic and modified signed log-likelihood ratio statistic for inference about the ratio of means of two independent lognormal distributions.

Krishnamoorthy and Mathew [34] developed exact confidence intervals and tests for the ratio (or the difference) of two lognormal means using the ideas of generalized $p$-values and generalized confidence intervals. Chen and Zhou [35] discussed interval estimates of five methods. These included the maximum likelihood approach, the bootstrap approach, the generalized approach and two methods based on the log-likelihood ratio statistic. They emphasize that parametric approaches were sometimes criticized because they did not perform well when assumptions were violated. Their own studies indicated that generalized approach and modified signed log-likelihood ratio approach were both fairly robust. Hanning et al. [36] introduced a subclass of generalized pivotal quantities and called fiducial generalized Pivotal Quantities (FGPQ) and showed that GCIs constructed using FGPQ have correct frequentist coverage under some mild conditions.

In recent years, various tests were developed for comparing the mean of two lognormal distributions. Li et al. [37] developed a new test by using the concept of fiducial and generalized $p$-value approach. Abdollahnezhad et al. [38] proposed a new test based on a generalized approach. Weng and Myers [39] assessed performance of confidence interval tests for the ratio of two lognormal means applied to Weibull and gamma distributed data. Jafari and Abdollahnezhad [40] developed a novel approach using CAT. More recently, Jiang et al. [41] proposed a higher-order likelihood-based method.

The generalized $p$-value has been widely used to handle the statistical testing problem involving nuisance parameters. We develop two simple procedures based on generalized approach. The developed methods are compared with the generalized confidence intervals by Weerahandi [12]. We also propose modified fiducial based approach using Fisher's fiducial inference for comparing means of two lognormal distributions. Li et al. [37] developed a fiducial based approach by borrowing the idea Fisher's fiducial inference. They focused on testing the equality of several normal means when the variances are unknown and unequal. A major motivation for the present work is to provide inference procedure for the small samples when the interest is to compare means of two lognormal distributions.

In Section 2, we introduce the BF problem and give confidence intervals based generalized approach. Section 3 presents five developed tests; the Z score test by Zhou et al. [32], the generalized $p$-value test by Krishnamoorthy and Mathew [34], the test based on generalized approach by Abdollahnezhad et al. [38], parametric bootstrap test and the modified fiducial based approach. Section 4 presents the simulation results to compare the methods. In section 5 we give two examples to illustrate the proposed approaches. Concluding remarks are given in Section 6.

## 2. The Behrens-Fisher (BF) problem

Assume that independent samples are available from two normal populations as $X_{i1},...,X_{in_i} iid N(\mu_i,\sigma_i^2), i=1,2$ where all four parameters $(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2)$ are unknown. Based on the above two independent samples, problem is to test

$$H_0:\mu_1=\mu_2 \quad \text{vs} \quad H_1:\mu_1\neq\mu_2 \tag{2.1}$$

First, we reduce the above data by sufficiency, and focus only on $\bar{X}_{i.}=\sum_{j=1}^{n_i} X_{ij}/n_i, S_i^2=\sum_{j=1}^{n_i}(X_{ij}-\bar{X}_{i.})^2, i=1,2$, where

$$\bar{X}_{i.} \sim N(\mu, \sigma_i^2/n_i) \quad \text{and} \quad S_i^2 \sim \sigma_i^2 \chi^2_{n_i-1} \quad , i = 1, 2 \tag{2.2}$$

and all four statistics are mutually independent. Let $\bar{x}_1, \bar{x}_2, s_1^2, s_2^2$ denote the observed values $\bar{X}_1, \bar{X}_2, S_1^2, S_2^2$, respectively. The inferences are to be based on the set of complete sufficient statistics whose distributions are given by equation 2.2. This problem is known as BF problem. We will construct generalized confidence interval for the parameter $\theta = \mu_1 - \mu_2$. We now will propose three generalized confidence intervals about the difference of the two normal means.

**2.1. Generalized Confidence Interval (GCI).** Weerahandi [12] defines a generalized pivotal as a statistic that has a distribution free of unknown parameters and an observed value that does not depend on nuisance parameters. The possibility of exact confidence interval can be achieved by extending the definition of confidence interval. The generalized pivotal is allowed to be a function of nuisance parameters. Weerahandi defines the confidence interval resulting from a generalized pivotal a generalized confidence interval. We shall now define generalized pivotal quantity for the difference between means of two normal population.

$$\begin{aligned} T(X_1, X_2; x_1, x_2, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2) &= \bar{x}_2 - \bar{x}_1 - \frac{\bar{X}_2 - \bar{X}_1 - (\mu_2 - \mu_1)}{\sqrt{\frac{\sigma_1^2}{n_1} \frac{\sigma_2^2}{n_2}}} \sqrt{\frac{\sigma_1^2 s_1^2}{n_1 S_1^2} + \frac{\sigma_2^2 s_2^2}{n_2 S_2^2}} \\ &= \bar{x}_2 - \bar{x}_1 - Z\sqrt{\frac{s_1^2}{U_1} + \frac{s_2^2}{U_2}} \end{aligned} \tag{2.3}$$

where

$$Z = \frac{\bar{X}_2 - \bar{X}_1 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1) \qquad , U_i = \frac{n_i S_i^2}{\sigma_i^2} \sim \chi^2_{n_i-1} \quad , i = 1, 2 \tag{2.4}$$

and $s_1, s_2, \bar{x}_1, \bar{x}_2$ are the observed values of $S_1, S_2, \bar{X}_1, \bar{X}_2$, respectively. Although it is difficult to obtain exact distribution of the test statistic defined in equation 2.3, we can approximate the distribution of this pivotal quantity by the Monte Carlo simulation study. The following algorithm can be used to approximate the distribution of 2.3.

**Algorithm 1**
For $i = 1$ to $m$
Generate values for $Z \sim N(0,1)$, $U_i \sim \chi^2_{(n_i-1)}$, $i = 1, 2$
Calculate $T$
(end $i$ loop)

Ordering the $m$ values of $T$ will provide us approximate distribution of the pivotal quantity.

If $T(1-\alpha)$ denotes the $100(1-\alpha)\%$th percentile of $T$, then $T(1-\alpha)$ is the $100(1-\alpha)\%$ generalized upper confidence interval for $\theta$. A $100(1-\alpha)\%$ generalized lower confidence interval for $\theta$ can be similarly obtained as $T(\alpha)$. A two-sided $100(1-\alpha)\%$ generalized confidence interval for $\theta$ is given by $(T(\alpha/2), T(1-\alpha/2))$.

**2.2. First approach based on generalized approach.** We will find a generalized pivotal quantity, $R$, using generalized approach. Since a generalized pivotal can be a function of all unknown parameters, we can construct $R$ based on the random quantities $Z_1 = (\sqrt{n_1}(\bar{X}_1 - \mu_1))/\sigma_1 \sim N(0,1)$, $Z_1 = (\sqrt{n_2}(\bar{X}_2 - \mu_1))/\sigma_1 \sim N(0,1)$, $U_1 = (n_1 S_1^2)/(\sigma_1^2 \sim \chi_(n_1 - 1)^2)$ and $U_2 = (n_2 S_2^2)/(\sigma_2^2 \sim \chi_(n_2 - 1)^2)$, whose distribution are free of unknown parameters. Using

$$\theta = (\bar{X}_2 - Z_2\sigma_2/\sqrt{n_2}) - (\bar{X}_1 - Z_1\sigma_1/\sqrt{n_1}) = (\bar{X}_2 - Z_2 S_2/\sqrt{U_2}) - (\bar{X}_1 - Z_1 S_1/\sqrt{U_1}) \tag{2.5}$$

we can define the following potential generalized pivotal;

$$\begin{aligned} R(X_1, X_2; x_1, x_2, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2) = & \quad (\bar{x}_2 - \bar{x}_1) - (Z_2 s_2/\sqrt{U_2} - Z_1 s_1/\sqrt{U_1}) \\ = & \quad (\bar{x}_2 - \bar{x}_1) - (T_2 s_2/\sqrt{n_2 - 1} - T_1 s_1/\sqrt{n_1 - 1}) \end{aligned} \tag{2.6}$$

where $s_1, s_2, \bar{x}, \bar{y}$ are the observed value of $S_1, S_2, \bar{X}, \bar{Y}$, respectively. Note that $T_1 \sim t(n_1-1)$ is independent of $T_2 \sim t(n_2-1)$ and $R(X, Y; x, y, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2) = \theta$.

If $R(1-\alpha)$ denotes the $100(1-\alpha)\%$th percentile of $R$, then $R(1-\alpha)$ is the $100(1-\alpha)\%$ generalized upper confidence interval for $\theta$. A $100(1-\alpha)\%$ generalized lower confidence interval for $\theta$ can be similarly obtained as $R(\alpha)$. A two-sided $100(1-\alpha)\%$ generalized confidence interval for $\theta$ is given by $(R(\alpha/2), R(1-\alpha/2))$. The approach can be summarized by the following algorithm.

**Algorithm 2**
For $i = 1$ to $m$
Generate values for $T_1 \sim t(n_1 - 1)$ and $T_2 \sim t(n_2 - 1)$
Calculate $R$
(end $i$ loop)

Order the $m$ values of $R$; find the $100\alpha$ and $100(1-\alpha)$ percentiles; denote these $R_{(l)}$ and $R_{(u)}$, respectively. A $100(1-\alpha)$ per cent confidence interval for $\theta$ is simply: $[R_{(l)}, R_{(u)}]$.

**2.3. Second approach based on generalized approach.** We will find a generalized pivotal quantity, $R^*$, using generalized approach.

$$\begin{aligned} R^*(X_1, X_2; x_1, x_2, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2) = & \quad \bar{x}_2 - \bar{x}_1 - \frac{\bar{X}_2 - \mu_2}{\sqrt{\frac{\sigma_2^2}{n_2}}}\sqrt{\frac{\sigma_2^2 s_2^2}{n_2 S_2^2}} + \frac{\bar{X}_1 - \mu_1}{\sqrt{\frac{\sigma_1^2}{n_1}}}\sqrt{\frac{\sigma_1^2 s_1^2}{n_1 S_1^2}} \\ = & \quad \bar{x}_2 - \bar{x}_1 - Z_2\sqrt{\frac{s_2^2}{U_2}} + Z_2\sqrt{\frac{s_2^2}{U_2}} \end{aligned} \tag{2.7}$$

where

$$(2.8)\qquad Z_i = \frac{\bar{X}_i - \mu_i}{\sqrt{\frac{\sigma_i^2}{n_i}}} \sim N(0,1) \qquad ,U_i = \frac{n_i S_i^2}{\sigma_i^2} \sim \chi^2_{n_i-1}, i = 1,2$$

If $R^*(1-\alpha)$ denotes the $100(1-\alpha)\%$th percentile of $R^*$, then $R^*(1-\alpha)$ is the $100(1-\alpha)\%$ generalized upper confidence interval for $\theta$. A $100(1-\alpha)\%$ generalized lower confidence interval for $\theta$ can be similarly obtained as $R^*(\alpha)$. A two-sided $100(1-\alpha)\%$ generalized confidence interval for $\theta$ is given by $(R^*(\alpha/2), R^*(1-\alpha/2))$. The approach can be summarized by the following algorithm.

**Algorithm 3**
For $i = 1$ to $m$
Generate values for $Z_i \sim N(0,1)$, $U_i \sim \chi^2_{n_i-1}\quad ,i = 1,2$
Calculate $R^*$
(end $i$ loop)

Order the $m$ values of $R^*$; find the $100\alpha$ and $100(1-\alpha)$ percentiles; denote these $R^*_{(l)}$ and $R^*_{(u)}$, respectively. A $100(1-\alpha)$ per cent confidence interval for $\theta$ is simply: $[R^*_{(l)}, R^*_{(u)}]$.

## 3. Comparing the means of two log-normal distributions

Let $X_1$ and $X_2$ be two independent lognormal random variables and $Y_1 = ln(X_1) \sim N(\mu_1, \sigma_1^2)$ and $Y_2 = ln(X_2) \sim N(\mu_2, \sigma_2^2)$. Considering that $\eta_1 = \mu_1 + (\sigma_1^2)/2$ and $\eta_2 = \mu_2 + (\sigma_2^2)/2$, it may be expressed as $E(X_1) = exp(\eta_1)$ and $E(X_2) = exp(\eta_2)$. Therefore the mean of two lognormal distributions can be reduced to inference related with $\eta_1 - \eta_2$ difference. Now let us handle hypotheses for $\eta_1 - \eta_2$.

Let $X_{1i}, i = 1, ..., n_1$ and $X_{2i}, i = 1, ..., n_2$ express two random samples from two independent lognormal distributions. Let $Y_{1i} = ln(X_{1i}), i = 1, ..., n_1$ and $Y_{2i} = ln(X_{2i}), i = 1, ..., n_2$ be two random variables that distribute normally. Accordingly

$$(3.1)\qquad \bar{Y}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} Y_{ij} \quad ,S_i^2 = \frac{1}{n_i - 1}\sum_{j=1}^{n_i}(Y_{ij} - \bar{Y}_i)^2, \quad i = 1,2$$

denotes the sample mean and sample variance of these two random variables, respectively. Let us hypotheses

$$(3.2)\qquad H_0 : \eta_1 \leq \eta_2 \quad \text{vs} \quad H_1 : \eta_1 > \eta_2$$

are considered.

We now will briefly introduce Z score test based on large sample test, two tests based on generalized $p$-value test, parametric bootstrap test and modified fiducial based approach.

**3.1. The Z test.** Zhou et al. [32] proposed a large sample test for hypotheses in 3.2. This test was expressed as;

$$Z = \frac{\bar{Y}_2 - \bar{Y}_1 + \frac{1}{2}(S_2^2 - S_1^2)}{\sqrt{S_1^2/n_1 + S_2^2/n_2 + \frac{1}{2}(S_1^4/(n_1-1) + S_2^2/(n_2-1))}} \tag{3.3}$$

The Z test have almost normal distribution for large samples considering that $H_0$ hypothesis is correct.

**3.2. The Krishnamoorthy and Mathew (KM) test.** Krishnamoorthy and Mathew [34] developed a new test for comparing the mean of two lognormal distributions using generalized $p$-value method. Let

$$T_{3i} = \bar{y}_i - \frac{Z_i}{U_i/\sqrt{n_i-1}}\frac{s_i}{\sqrt{n_i}} + \frac{1}{2}\frac{s_i^2}{U_i^2/(n_i-1)} \quad ,i=1,2 \tag{3.4}$$

where $Z_i = \frac{\sqrt{n_i}(\bar{Y}_i-\mu_i)}{\sigma_i} \sim N(0,1), i=1,2$ and $U_i^2 = (n_i-1)S_i^2/\sigma_i^2 \sim \chi^2_{n_i-1}, i=1,2$ are independent random variables. Let

$$T_3 = T_{31} - T_{32} - (\eta_1 - \eta_2) \tag{3.5}$$

is the generalized test variables and the generalized $p$-value can be computed as;

$$p = P(T_3 \leq 0|\eta_1 - \eta_2 = 0) \tag{3.6}$$

**3.3. The test by Abdollahnezhad et al. (AB).** Abdollahnezhad et al. [38] proposed a generalized approach to obtain the $p$-value for hypotheses in 3.2. The MLE's for $\mu_i$ and $\sigma_i^2 (i=1,2)$ are $\bar{Y}_i$ and $S_i^2$, respectively, where

$$\bar{Y}_i = \frac{1}{n_i}\sum_{i=1}^{n} Y_{ij} \quad ,S_i^2 = \frac{1}{n_i}\sum_{i=1}^{n}(Y_{ij}-\bar{Y}_i)^2 \tag{3.7}$$

A generalized variable uses the generalized p-value concept, expressed as;

$$\begin{aligned} T = \quad & \bar{y}_{1.} - \bar{y}_{2.} + \frac{\bar{Y}_{2.}-\bar{Y}_{1.}-(\mu_2-\mu_1)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}\sqrt{\frac{\sigma_1^2 s_1^2}{n_1 S_1^2} + \frac{\sigma_2^2 s_1^2}{n_2 S_2^2}} + \frac{\sigma_1^2 s_1^2}{2S_1^2} - \frac{\sigma_2^2 s_1^2}{2S_2^2} - (\eta_1-\eta_2) \\ = \quad & \bar{y}_{1.} - \bar{y}_{2.} + Z\sqrt{\frac{s_1^2}{u_1}+\frac{s_2^2}{u_2}} + \frac{n_1 s_1^2}{2U_1} - \frac{n_2 s_2^2}{2U_2} - (\eta_1-\eta_2) \end{aligned} \tag{3.8}$$

where

$$Z = \frac{\bar{Y}_{2.} - \bar{Y}_{1.} - (\mu_2-\mu_1)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \sim N(0,1) \quad ,U_i = \frac{n_i S_i^2}{\sigma_i^2} \sim \chi^2_{(n_i-1)} \quad ,i=1,2 \tag{3.9}$$

and, $(\bar{y}_1, \bar{y}_2)$ and $(s_1^2, s_2^2)$ are observed values of $(\bar{Y}_1, \bar{Y}_2)$ and $(S_1^2, S_2^2)$, respectively. Thus the generalized $p$-value for the null hypothesis in 3.1 is given by

$$(3.10) \quad p = P(T \leq t_{obs} | \eta_1 - \eta_2 = 0) = E\left(\Phi\left(\frac{\bar{y}_{2.} - \bar{y}_{1.} + \frac{n_2 s_2^2}{2U_2} - \frac{n_1 s_1^2}{2U_1}}{\sqrt{\frac{s_1^2}{U_1} + \frac{s_2^2}{u_2}}}\right)\right)$$

where $\Phi(.)$ is the standard normal distribution function and the expectation is taken with respect to independent chi-square random variables, $U_1$ and $U_2$.

**3.4. The parametric bootstrap (PB) test.** The Parametric bootstrap (PB) is widely used in many statistical inferential procedures. The PB approach is a type of Monte Carlo method which can be applied in situations where samples or sample statistics are not easy to derive. Krishnamoorthy et al. [42] used the PB method to test the equality of several means when the variances are unknown and arbitrary.

Let us consider hypotheses in 3.2. It is well known that $Y_i$ and $S_i^2$ are independent and

$$(3.11) \quad Y_i \sim N(\mu_i, \sigma_i^2/n_i), \qquad \frac{(n_i - 1)S_i^2}{\sigma_i^2} \sim \chi^2_{(n_i-1)} \quad, i = 1, 2$$

The natural (unbiased) estimators of $\eta_1$ and $\eta_2$ are

$$(3.12) \quad \hat{\eta}_1 = \bar{Y}_1 + \frac{S_1^2}{2} \quad \text{and} \quad \hat{\eta}_2 = \bar{Y}_2 + \frac{S_2^2}{2}$$

Note that

$$(3.13) \quad var(\hat{\eta}_1 - \hat{\eta}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} + \frac{1}{2}\left(\frac{\sigma_1^4}{(n_1 - 1)} + \frac{\sigma_2^4}{(n_2 - 1)}\right)$$

and an unbiased estimator of 3.13 is

$$(3.14) \quad var\left(\bar{Y}_1 - \bar{Y}_2 + \frac{S_1^2}{2} - \frac{S_2^2}{2}\right) = \frac{S_1^2}{n_1} + \frac{S_2^2}{n_2} + \frac{1}{2}\left(\frac{S_1^4}{n_1 + 1} + \frac{S_2^4}{n_2 + 1}\right)$$

under the null hypothesis in 3.1 the natural test statistic,

$$(3.15) \quad T = \frac{(\hat{\eta}_1 - \hat{\eta}_2)}{\sqrt{var(\hat{\eta}_1 - \hat{\eta}_2)}} = \frac{\bar{Y}_1 - \bar{Y}_2 + \frac{S_1^2}{2} - \frac{S_2^2}{2}}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2} + \frac{1}{2}\left(\frac{S_1^4}{n_1+1} + \frac{S_2^4}{n_2+1}\right)}}$$

To find approximate the distribution of $T$ in Equation 3.15 using a PB approach, we can take common mean to be zero. The parametric bootstrap pivot variable can be proposed as follows. The PB pivot variable based on the test statistic in Equation 3.15 is given by

$$(3.16)\quad T_{PB} = \frac{(\hat{\eta}_1 - \hat{\eta}_2) - (\eta_1 - \eta_2)}{\sqrt{var(\hat{\eta}_1 - \hat{\eta}_2)}} = \frac{\bar{Y}_{B1} - \bar{Y}_{B2} + \frac{1}{2}(S^2_{B1} - S^2_{B2}) - \frac{1}{2}(s^2_1 - s^2_2)}{\sqrt{var(\hat{\eta}_1 - \hat{\eta}_2)}}$$

where

$$(3.17)\quad \bar{Y}_{Bi} \sim N\left(0, \frac{S^2_i}{n_i}\right), \quad S^2_{Bi} \sim \frac{S^2_i \chi^2_{(n_i-1)}}{n_i - 1}$$

$$(3.18)\quad var_B(\hat{\eta}_1 - \hat{\eta}_2) = \frac{S^2_{B1}}{n_1} + \frac{S^2_{B2}}{n_2} + \frac{1}{2}\left(\frac{(S^2_{B1})^2}{(n_1+1)} + \frac{(S^2_{B2})^2}{(n_2+1)}\right) \quad ,i = 1,2$$

Given $s^2_1$ and $s^2_2$ are the observed values of $S^2_1$ and $S^2_2$, the PB $p$-value can be computed as,

$$(3.19)\quad P(T_{PB}(s^2_i; \chi^2_{(n_i-1)}) > t_{PB}) < \alpha \quad ,i = 1,2$$

where $t_{PB}$ is the observed of $T$ in Equation 3.15. The above probability does not depend on any unknown parameters and so it can be estimated using Monte Carlo simulation given in the following algorithm.

**Algorithm 4**
For a given $(n_1, n_2), (\bar{x}_1, \bar{x}_2)$ and $(s^2_1, s^2_2)$
Compute $T$ in (18) and call it $t_{PB}$
For $j = 1, m$
Compute $T_{PB}$ in 3.16
If $T_{PB} > t_{PB}$ set $K_j = 1$
(end loop)
$(1/m)\sum_{j=1}^{m} K_j$ is a Monte-Carlo estimate of the $PB$ $p$-value in 3.19.

**3.5. The modified fiducial based (MFB) test.** In this section, the fiducial approach of Fisher [43] and generalized $p$-value approach are combined and modification of KM test is proposed for the means of two lognormal distributions.

Let $U_{1i} \sim N(0,1)$ and $U_{2i} \sim \chi_(n_i - 1)^2, i = 1,2$ be two different independent variables. It is known that $\bar{Y}_i \sim N(\mu_i, (\sigma^2_i)/\sqrt{n_i})$ and $(n_i - 1)S^2_i/\sigma^2_i \sim \chi_(n_i - 1)^2$. Let us express $Y_i$ and $(n_i - 1)S^2_i$ expressions as the functions of $U_{1i}$ and $U_{2i}$;

$$(3.20)\quad \bar{Y}_i = \mu_i + \frac{\sigma_i}{\sqrt{n_i}}, (n_i - 1)S^2_i = \sigma^2_i U_{2i} \quad ,i = 1,2$$

when $(\bar{y}_i, s^2_i)$ and $(u_{1i}, u_{2i})$ observation values are given, equations in 3.20 can be written in the form of $\bar{y}_i = \mu_i + \frac{\sigma^2_i}{n_i} u_{1i}$ and $(n_i - 1)s^2_i = \sigma^2_i u_{2i}$. From these two equations

$$(3.21)\quad \mu_i = y_i - \frac{u_{1i}}{\sqrt{u_{2i}}/(n_i - 1)}\sqrt{\frac{s^2_i}{n_i}} \quad ,\sigma^2_i = \frac{(n_i - 1)s^2_i}{u_{2i}}$$

when $(\bar{y}_i, s_i^2)$ is given the fiducial test statistics of $\mu_i$ and $\sigma_i^2$;

$$(3.22)\quad T_{\mu_i} = \bar{y}_i - t_i\sqrt{s_i^2/n_i} \quad , T_{\sigma_i^2} = \frac{(n_i-1)s_i^2}{2(q_i)}$$

where $t_i \sim t(n_i - 1)$ and $q_i \sim \sigma^2_{n_i-1} \quad , i = 1, 2$. Let

$$(3.23)\quad T_i = \bar{y}_i - t_i\sqrt{s_i^2/n_i} + (n_i - 1)s_i^2/2(q_i) \quad , i = 1, 2$$

The test statistics becomes;

$$(3.24)\quad T_F = T_1 - T_2$$

Accordingly the fiducial $p$-value depending on the fiducial approach for the hypotheses in 3.2 becomes;

$$(3.25)\quad p = P(T_F \leq 0 | \eta_1 - \eta_2 = 0)$$

## 4. Simulation study

In this section, A Monte Carlo simulation study is conducted to evaluate performances of the proposed methods under different scenarios. The three proposed confidence intervals for the BF problem are compared to evaluate expected lengths and coverage probabilities. The nominal level of the confidence intervals is 95%. Two configuration factors were taken into account to evaluate the performance of confidence intervals; sample size and variance. To obtain two generalized confidence intervals, we used a two-step simulation. For each step, we used simulation consisting of 2000 runs.

A Monte Carlo simulation study also is performed on the proposed methods in order to compare and evaluate them in terms of type I error probabilities and powers for the nominal value $\alpha = 0.05$. To estimate the type I error rates and power of Z test, we used simulation consisting of 100,000 runs for each of the sample size and parameter configurations. To estimate the type I error rates and power of the KM, AB, PB and MFB tests, we have used a two-step simulation. For a given sample size and parameter configuration, we generated 2500 observed vectors $(\bar{x}_1, \bar{x}_2, s_1^2, s_2^2)$ and used 5000 runs to estimate the p-value in Equations 3.6, 3.10, 3.19 and 3.25.

Table 1 shows that the generalized confidence interval based on equation 2.7 has the good expected length when sample sizes are equal or large. However, its coverage probabilities perform very poorly, especially when the sample sizes are unequal. The expected lengths of generalized confidence interval based on equation 2.6 are shorter than GCI method. Its coverage probabilities have satisfactory, especially when the simple sizes are moderate and large. The coverage probability of the GCI method performs second.

Table 2 and Table 3 illustrate the type I errors of the proposed tests for comparing the mean of two lognormal distributions when variances are equal and unequal respectively. Table 2 and Table 3 show that,

1. The type I error rates of the Z test seems to be very conservative when equal and small sample sizes are associated with large variances. However, its type I error rate is very close to nominal level for large samples. It is noted that Z test is either too conservative or too liberal when the sample sizes are different.
2. Results reveal that AB test performs extremely poor in respect to its type I error rates when the heteroscedasticity is present.
3. It is noted that the KM and MFP tests are performing satisfactory in respect to their type I error rates and results close to each other in all cases.
4. For the small samples and unequal sample sizes, PB overestimates the type I error rate.

The powers of the five tests are presented in Table 3. These result shows that the power of the AB test is smallest among the five tests. The MFB test appears to be more powerful than the other tests in most cases. As expected, for $n_1$ and $n_2$ large ($> 25$), the powers of the Z, KM, PB and MFB tests are nearly identical. The PB test is the second best after MFB test in most case.

**Table 1:** Expected length (EL) and coverage probability (CP) of the proposed methods for $\theta = \mu_1 - \mu_2$ with nominal 95%.

| $(n_1, n_2)$ | $(\sigma_1^2, \sigma_2^2)$ | Algorithm 1 | | Algorithm 2 | | GCI | |
|---|---|---|---|---|---|---|---|
| | | EL | CP | EL | CP | EL | CP |
| (5,5) | (1,1) | 3.822 | 0.988 | 3.828 | 0.985 | 3.823 | 0.988 |
| | (1,5) | 6.503 | 0.980 | 6.520 | 0.981 | 6.509 | 0.980 |
| | (1,10) | 8.734 | 0.975 | 8.754 | 0.975 | 8.740 | 0.973 |
| (10,10) | (1,1) | 2.096 | 0.970 | 2.096 | 0.975 | 2.097 | 0.971 |
| | (1,5) | 3.606 | 0.967 | 3.606 | 0.969 | 3.610 | 0.967 |
| | (1,10) | 4.870 | 0.964 | 4.869 | 0.966 | 4.874 | 0.965 |
| (10,20) | (1,1) | 1.756 | 0.968 | 2.102 | 0.990 | 1.759 | 0.969 |
| | (1,5) | 2.591 | 0.966 | 3.636 | 0.996 | 2.589 | 0.969 |
| | (1,10) | 3.348 | 0.963 | 4.920 | 0.998 | 3.348 | 0.964 |
| (20,10) | (1,1) | 1.754 | 0.969 | 1.340 | 0.905 | 1.758 | 0.966 |
| | (1,5) | 3.419 | 0.963 | 2.301 | 0.859 | 3.426 | 0.962 |
| | (1,10) | 4.738 | 0.960 | 3.105 | 0.846 | 4.748 | 0.960 |
| (25,25) | (1,1) | 1.181 | 0.953 | 1.182 | 0.956 | 1.184 | 0.960 |
| | (1,5) | 2.042 | 0.957 | 2.041 | 0.957 | 2.047 | 0.954 |
| | (1,10) | 2.760 | 0.953 | 2.761 | 0.954 | 2.769 | 0.954 |
| (25,50) | (1,1) | 1.013 | 0.957 | 1.183 | 0.982 | 1.014 | 0.956 |
| | (1,5) | 1.524 | 0.950 | 2.049 | 0.993 | 1.528 | 0.954 |
| | (1,10) | 1.987 | 0.952 | 2.773 | 0.994 | 1.991 | 0.949 |
| (50,25) | (1,1) | 1.011 | 0.955 | 0.807 | 0.900 | 1.013 | 0.958 |
| | (1,5) | 1.949 | 0.955 | 1.392 | 0.867 | 1.951 | 0.954 |
| | (1,10) | 2.697 | 0.949 | 1.882 | 0.854 | 2.699 | 0.950 |
| (50,50) | (1,1) | 0.808 | 0.950 | 0.808 | 0.953 | 0.809 | 0.951 |
| | (1,5) | 1.398 | 0.950 | 1.398 | 0.952 | 1.399 | 0.949 |
| | (1,10) | 1.891 | 0.950 | 1.890 | 0.950 | 1.893 | 0.947 |
| (50,100) | (1,1) | 0.698 | 0.954 | 0.809 | 0.976 | 0.698 | 0.954 |
| | (1,5) | 1.058 | 0.951 | 1.400 | 0.991 | 1.058 | 0.952 |
| | (1,10) | 1.381 | 0.949 | 1.894 | 0.993 | 1.382 | 0.950 |
| (100,50) | (1,1) | 0.695 | 0.948 | 0.562 | 0.890 | 0.696 | 0.951 |
| | (1,5) | 1.335 | 0.955 | 0.971 | 0.868 | 1.335 | 0.949 |
| | (1,10) | 1.845 | 0.952 | 1.314 | 0.853 | 1.846 | 0.948 |
| (100,100) | (1,1) | 0.562 | 0.952 | 0.562 | 0.953 | 0.562 | 0.954 |
| | (1,5) | 0.972 | 0.951 | 0.973 | 0.951 | 0.974 | 0.953 |
| | (1,10) | 1.316 | 0.948 | 1.317 | 0.948 | 1.318 | 0.948 |
| (100,200) | (1,1) | 0.486 | 0.954 | 0.561 | 0.977 | 0.487 | 0.955 |
| | (1,5) | 0.740 | 0.950 | 0.974 | 0.990 | 0.741 | 0.950 |
| | (1,10) | 0.968 | 0.951 | 1.319 | 0.992 | 0.969 | 0.951 |
| (200,100) | (1,1) | 0.485 | 0.953 | 0.394 | 0.883 | 0.486 | 0.956 |
| | (1,5) | 0.930 | 0.950 | 0.682 | 0.851 | 0.931 | 0.948 |
| | (1,10) | 1.286 | 0.946 | 0.922 | 0.849 | 1.287 | 0.946 |
| (200,200) | (1,1) | 0.395 | 0.951 | 0.394 | 0.952 | 0.395 | 0.952 |
| | (1,5) | 0.683 | 0.949 | 0.683 | 0.950 | 0.684 | 0.949 |
| | (1,10) | 0.925 | 0.947 | 0.924 | 0.946 | 0.926 | 0.947 |

**Table 2:** Type I error probabilities for five tests at 5% significance level when $H_0 : \eta_1 \leq \eta_2$ vs $H_1 : \eta_1 > \eta_2$ and $(\mu_1, \mu_2) = (0, 0)$.

| $(\sigma_1^2, \sigma_2^2)$ | Z | KM | AB | PB | MFB | Z | KM | AB | PB | MFB |
|---|---|---|---|---|---|---|---|---|---|---|
| | $(n_1, n_2) = (5, 5)$ | | | | | $(n_1, n_2) = (20, 10)$ | | | | |
| (2,2) | 0.049 | 0.046 | 0.042 | 0.091 | 0.032 | 0.074 | 0.049 | 0.046 | 0.072 | 0.043 |
| (4,4) | 0.032 | 0.046 | 0.044 | 0.124 | 0.039 | 0.075 | 0.048 | 0.045 | 0.090 | 0.041 |
| (6,6) | 0.023 | 0.046 | 0.046 | 0.130 | 0.041 | 0.076 | 0.046 | 0.044 | 0.091 | 0.042 |
| (8,8) | 0.019 | 0.048 | 0.046 | 0.136 | 0.047 | 0.076 | 0.044 | 0.043 | 0.091 | 0.045 |
| (10,10) | 0.015 | 0.047 | 0.045 | 0.142 | 0.048 | 0.077 | 0.045 | 0.044 | 0.092 | 0.047 |
| | $(n_1, n_2) = (10, 10)$ | | | | | $(n_1, n_2) = (25, 50)$ | | | | |
| (2,2) | 0.043 | 0.044 | 0.046 | 0.069 | 0.033 | 0.050 | 0.047 | 0.054 | 0.060 | 0.045 |
| (4,4) | 0.032 | 0.044 | 0.045 | 0.068 | 0.042 | 0.046 | 0.047 | 0.059 | 0.059 | 0.041 |
| (6,6) | 0.025 | 0.039 | 0.048 | 0.077 | 0.046 | 0.046 | 0.046 | 0.055 | 0.060 | 0.043 |
| (8,8) | 0.020 | 0.040 | 0.048 | 0.079 | 0.048 | 0.044 | 0.048 | 0.054 | 0.062 | 0.043 |
| (10,10) | 0.018 | 0.041 | 0.048 | 0.081 | 0.047 | 0.044 | 0.048 | 0.053 | 0.063 | 0.043 |
| | $(n_1, n_2) = (25, 25)$ | | | | | $(n_1, n_2) = (50, 25)$ | | | | |
| (2,2) | 0.045 | 0.041 | 0.045 | 0.059 | 0.037 | 0.066 | 0.046 | 0.057 | 0.068 | 0.042 |
| (4,4) | 0.041 | 0.049 | 0.049 | 0.060 | 0.044 | 0.069 | 0.051 | 0.058 | 0.066 | 0.046 |
| (6,6) | 0.037 | 0.045 | 0.049 | 0.064 | 0.050 | 0.069 | 0.051 | 0.058 | 0.069 | 0.049 |
| (8,8) | 0.035 | 0.045 | 0.047 | 0.065 | 0.050 | 0.070 | 0.052 | 0.058 | 0.072 | 0.052 |
| (10,10) | 0.034 | 0.046 | 0.048 | 0.064 | 0.050 | 0.070 | 0.052 | 0.059 | 0.071 | 0.051 |
| | $(n_1, n_2) = (50, 50)$ | | | | | $(n_1, n_2) = (25, 100)$ | | | | |
| (2,2) | 0.047 | 0.043 | 0.045 | 0.055 | 0.044 | 0.028 | 0.058 | 0.055 | 0.051 | 0.046 |
| (4,4) | 0.045 | 0.043 | 0.042 | 0.055 | 0.043 | 0.021 | 0.051 | 0.051 | 0.053 | 0.046 |
| (6,6) | 0.042 | 0.044 | 0.045 | 0.053 | 0.044 | 0.018 | 0.050 | 0.050 | 0.048 | 0.042 |
| (8,8) | 0.043 | 0.043 | 0.046 | 0.055 | 0.045 | 0.016 | 0.047 | 0.052 | 0.046 | 0.040 |
| (10,10) | 0.042 | 0.046 | 0.047 | 0.056 | 0.045 | 0.015 | 0.045 | 0.053 | 0.051 | 0.040 |
| | $(n_1, n_2) = (100, 100)$ | | | | | | | | | |
| (2,2) | 0.049 | 0.046 | 0.042 | 0.052 | 0.044 | 0.056 | 0.046 | 0.040 | 0.051 | 0.051 |
| (4,4) | 0.048 | 0.046 | 0.049 | 0.056 | 0.041 | 0.056 | 0.052 | 0.039 | 0.053 | 0.058 |
| (6,6) | 0.047 | 0.046 | 0.050 | 0.054 | 0.043 | 0.057 | 0.050 | 0.038 | 0.048 | 0.055 |
| (8,8) | 0.045 | 0.043 | 0.050 | 0.057 | 0.044 | 0.059 | 0.049 | 0.040 | 0.046 | 0.055 |
| (10,10) | 0.046 | 0.042 | 0.049 | 0.060 | 0.045 | 0.058 | 0.049 | 0.040 | 0.051 | 0.056 |
| | $(n_1, n_2) = (200, 200)$ | | | | | | | | | |
| (2,2) | 0.049 | 0.046 | 0.052 | 0.056 | 0.044 | 0.060 | 0.042 | 0.051 | 0.059 | 0.045 |
| (4,4) | 0.047 | 0.046 | 0.052 | 0.050 | 0.041 | 0.064 | 0.046 | 0.052 | 0.054 | 0.049 |
| (6,6) | 0.050 | 0.044 | 0.054 | 0.052 | 0.045 | 0.065 | 0.045 | 0.049 | 0.056 | 0.044 |
| (8,8) | 0.048 | 0.044 | 0.056 | 0.055 | 0.046 | 0.066 | 0.045 | 0.050 | 0.062 | 0.047 |
| (10,10) | 0.049 | 0.044 | 0.052 | 0.053 | 0.047 | 0.066 | 0.049 | 0.050 | 0.062 | 0.048 |
| | $(n_1, n_2) = (10, 20)$ | | | | | | | | | |
| (2,2) | 0.051 | 0.043 | 0.047 | 0.054 | 0.040 | 0.050 | 0.049 | 0.041 | 0.047 | 0.048 |
| (4,4) | 0.045 | 0.044 | 0.051 | 0.056 | 0.040 | 0.047 | 0.050 | 0.041 | 0.051 | 0.044 |
| (6,6) | 0.043 | 0.042 | 0.050 | 0.059 | 0.042 | 0.048 | 0.052 | 0.043 | 0.054 | 0.046 |
| (8,8) | 0.042 | 0.044 | 0.050 | 0.059 | 0.041 | 0.046 | 0.052 | 0.043 | 0.054 | 0.045 |
| (10,10) | 0.041 | 0.045 | 0.051 | 0.059 | 0.040 | 0.048 | 0.049 | 0.045 | 0.054 | 0.045 |

**Table 3:** Type I error probabilities for five tests at 5% significance level when $H_0 : \eta_1 \leq \eta_2$ vs $H_1 : \eta_1 > \eta_2$.

| $(\mu_1, \mu_2)$ | $(\sigma_1^2, \sigma_2^2)$ | Z | KM | AB | PB | MFB | Z | KM | AB | PB | MFB |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | $(n_1, n_2) = (5, 5)$ | | | | | $(n_1, n_2) = (10, 25)$ | | | | |
| (1,0) | (2,4) | 0.101 | 0.056 | 0.097 | 0.128 | 0.040 | 0.047 | 0.050 | 0.294 | 0.055 | 0.042 |
| (1,0) | (6,8) | 0.053 | 0.055 | 0.091 | 0.148 | 0.043 | 0.020 | 0.048 | 0.243 | 0.061 | 0.036 |
| (1,0) | (12,14) | 0.029 | 0.053 | 0.086 | 0.119 | 0.045 | 0.010 | 0.048 | 0.198 | 0.055 | 0.042 |
| (2,0) | (2,6) | 0.142 | 0.061 | 0.173 | 0.124 | 0.050 | 0.068 | 0.056 | 0.626 | 0.066 | 0.053 |
| (2,0) | (10,14) | 0.053 | 0.058 | 0.141 | 0.132 | 0.051 | 0.018 | 0.049 | 0.472 | 0.066 | 0.044 |
| (2,0) | (16,20) | 0.033 | 0.057 | 0.129 | 0.128 | 0.051 | 0.011 | 0.047 | 0.401 | 0.058 | 0.044 |
| | | $(n_1, n_2) = (10, 10)$ | | | | | $(n_1, n_2) = (50, 25)$ | | | | |
| (1,0) | (2,4) | 0.089 | 0.056 | 0.168 | 0.091 | 0.049 | 0.090 | 0.051 | 0.362 | 0.070 | 0.057 |
| (1,0) | (6,8) | 0.059 | 0.047 | 0.147 | 0.109 | 0.044 | 0.083 | 0.053 | 0.304 | 0.072 | 0.055 |
| (1,0) | (12,14) | 0.042 | 0.048 | 0.125 | 0.090 | 0.044 | 0.080 | 0.052 | 0.255 | 0.073 | 0.054 |
| (2,0) | (2,6) | 0.115 | 0.052 | 0.330 | 0.086 | 0.055 | 0.102 | 0.056 | 0.724 | 0.066 | 0.059 |
| (2,0) | (10,14) | 0.062 | 0.054 | 0.259 | 0.101 | 0.049 | 0.088 | 0.052 | 0.569 | 0.075 | 0.052 |
| (2,0) | (16,20) | 0.048 | 0.052 | 0.226 | 0.094 | 0.048 | 0.084 | 0.054 | 0.489 | 0.076 | 0.049 |
| | | $(n_1, n_2) = (25, 25)$ | | | | | $(n_1, n_2) = (100, 25)$ | | | | |
| (1,0) | (2,4) | 0.076 | 0.054 | 0.332 | 0.069 | 0.056 | 0.097 | 0.052 | 0.372 | 0.052 | 0.058 |
| (1,0) | (6,8) | 0.060 | 0.051 | 0.284 | 0.077 | 0.050 | 0.097 | 0.052 | 0.292 | 0.055 | 0.055 |
| (1,0) | (12,14) | 0.052 | 0.050 | 0.227 | 0.069 | 0.047 | 0.097 | 0.049 | 0.232 | 0.062 | 0.054 |
| (2,0) | (2,6) | 0.092 | 0.055 | 0.694 | 0.065 | 0.055 | 0.106 | 0.057 | 0.747 | 0.054 | 0.056 |
| (2,0) | (10,14) | 0.064 | 0.055 | 0.511 | 0.065 | 0.049 | 0.102 | 0.051 | 0.585 | 0.059 | 0.053 |
| (2,0) | (16,20) | 0.056 | 0.053 | 0.443 | 0.072 | 0.049 | 0.101 | 0.051 | 0.509 | 0.061 | 0.055 |
| | | $(n_1, n_2) = (50, 50)$ | | | | | $(n_1, n_2) = (10, 100)$ | | | | |
| (1,0) | (2,4) | 0.068 | 0.054 | 0.549 | 0.053 | 0.057 | 0.019 | 0.060 | 0.475 | 0.053 | 0.036 |
| (1,0) | (6,8) | 0.059 | 0.047 | 0.455 | 0.057 | 0.048 | 0.003 | 0.053 | 0.359 | 0.042 | 0.038 |
| (1,0) | (12,14) | 0.053 | 0.051 | 0.346 | 0.053 | 0.049 | 0.001 | 0.045 | 0.282 | 0.037 | 0.041 |
| (2,0) | (2,6) | 0.080 | 0.055 | 0.945 | 0.048 | 0.055 | 0.028 | 0.056 | 0.930 | 0.055 | 0.043 |
| (2,0) | (10,14) | 0.061 | 0.051 | 0.799 | 0.054 | 0.052 | 0.002 | 0.046 | 0.702 | 0.042 | 0.039 |
| (2,0) | (16,20) | 0.056 | 0.052 | 0.713 | 0.055 | 0.050 | 0.001 | 0.045 | 0.595 | 0.033 | 0.039 |
| | | $(n_1, n_2) = (100, 100)$ | | | | | $(n_1, n_2) = (25, 100)$ | | | | |
| (1,0) | (2,4) | 0.063 | 0.051 | 0.836 | 0.043 | 0.053 | 0.040 | 0.049 | 0.621 | 0.048 | 0.048 |
| (1,0) | (6,8) | 0.057 | 0.047 | 0.712 | 0.052 | 0.045 | 0.024 | 0.046 | 0.478 | 0.045 | 0.045 |
| (1,0) | (12,14) | 0.053 | 0.046 | 0.567 | 0.059 | 0.047 | 0.017 | 0.048 | 0.377 | 0.042 | 0.042 |
| (2,0) | (2,6) | 0.072 | 0.054 | 0.997 | 0.051 | 0.049 | 0.052 | 0.058 | 0.989 | 0.056 | 0.056 |
| (2,0) | (10,14) | 0.059 | 0.048 | 0.980 | 0.053 | 0.049 | 0.023 | 0.052 | 0.893 | 0.046 | 0.046 |
| (2,0) | (16,20) | 0.055 | 0.048 | 0.941 | 0.059 | 0.047 | 0.018 | 0.048 | 0.801 | 0.048 | 0.048 |
| | | $(n_1, n_2) = (5, 25)$ | | | | | $(n_1, n_2) = (50, 100)$ | | | | |
| (1,0) | (2,4) | 0.025 | 0.050 | 0.264 | 0.054 | 0.030 | 0.054 | 0.051 | 0.758 | 0.047 | 0.049 |
| (1,0) | (6,8) | 0.004 | 0.050 | 0.218 | 0.051 | 0.029 | 0.042 | 0.048 | 0.623 | 0.054 | 0.045 |
| (1,0) | (12,14) | 0.001 | 0.049 | 0.176 | 0.040 | 0.031 | 0.037 | 0.049 | 0.486 | 0.058 | 0.045 |
| (2,0) | (2,6) | 0.040 | 0.061 | 0.590 | 0.066 | 0.038 | 0.064 | 0.055 | 0.996 | 0.044 | 0.056 |
| (2,0) | (10,14) | 0.002 | 0.051 | 0.405 | 0.052 | 0.034 | 0.043 | 0.047 | 0.949 | 0.056 | 0.048 |
| (2,0) | (16,20) | 0.001 | 0.051 | 0.354 | 0.039 | 0.037 | 0.038 | 0.049 | 0.893 | 0.058 | 0.048 |

**Table 4:** Powers of five tests for $H_0 : \eta_1 \leq \eta_2$ vs $H_1 : \eta_1 > \eta_2$ hypothesis and $(\mu_1, \mu_2) = (0, 0)$.

| $(\sigma_1^2, \sigma_2^2)$ | Z | KM | AB | PB | MFB | Z | KM | AB | PB | MFB |
|---|---|---|---|---|---|---|---|---|---|---|
| | $(n_1, n_2) = (5, 5)$ | | | | | $(n_1, n_2) = (50, 50)$ | | | | |
| (4,2) | 0.060 | 0.106 | 0.067 | 0.240 | 0.114 | 0.547 | 0.592 | 0.206 | 0.606 | 0.581 |
| (6,2) | 0.069 | 0.166 | 0.086 | 0.367 | 0.208 | 0.890 | 0.915 | 0.397 | 0.917 | 0.910 |
| (8,2) | 0.078 | 0.232 | 0.100 | 0.482 | 0.270 | 0.976 | 0.987 | 0.546 | 0.989 | 0.986 |
| (10,2) | 0.082 | 0.354 | 0.116 | 0.557 | 0.390 | 0.994 | 1.00 | 0.667 | 0.998 | 1.00 |
| (12,2) | 0.085 | 0.380 | 0.129 | 0.620 | 0.422 | 0.999 | 1.00 | 0.742 | 1.00 | 1.00 |
| | $(n_1, n_2) = (10, 10)$ | | | | | $(n_1, n_2) = (25, 100)$ | | | | |
| (4,2) | 0.125 | 0.190 | 0.085 | 0.285 | 0.226 | 0.350 | 0.506 | 0.197 | 0.561 | 0.474 |
| (6,2) | 0.219 | 0.352 | 0.126 | 0.472 | 0.458 | 0.674 | 0.810 | 0.336 | 0.903 | 0.785 |
| (8,2) | 0.311 | 0.498 | 0.171 | 0.621 | 0.642 | 0.843 | 0.930 | 0.448 | 0.989 | 0.921 |
| (10,2) | 0.394 | 0.602 | 0.200 | 0.714 | 0.744 | 0.921 | 0.969 | 0.545 | 0.992 | 0.967 |
| (12,2) | 0.462 | 0.668 | 0.249 | 0.781 | 0.794 | 0.958 | 0.986 | 0.611 | 1.00 | 0.984 |
| | $(n_1, n_2) = (10, 20)$ | | | | | $(n_1, n_2) = (100, 25)$ | | | | |
| (4,2) | 0.111 | 0.256 | 0.109 | 0.267 | 0.328 | 0.593 | 0.512 | 0.148 | 0.564 | 0.644 |
| (6,2) | 0.218 | 0.446 | 0.170 | 0.459 | 0.592 | 0.928 | 0.880 | 0.323 | 0.914 | 0.948 |
| (8,2) | 0.317 | 0.610 | 0.226 | 0.596 | 0.772 | 0.991 | 0.982 | 0.467 | 0.991 | 0.996 |
| (10,2) | 0.405 | 0.706 | 0.281 | 0.709 | 0.824 | 0.999 | 1.00 | 0.570 | 1.00 | 1.00 |
| (12,2) | 0.477 | 0.762 | 0.321 | 0.779 | 0.916 | 1.00 | 1.00 | 0.650 | 1.00 | 1.00 |
| | $(n_1, n_2) = (25, 25)$ | | | | | $(n_1, n_2) = (50, 100)$ | | | | |
| (4,2) | 0.312 | 0.360 | 0.134 | 0.390 | 0.522 | 0.602 | 0.692 | 0.240 | 0.676 | 0.882 |
| (6,2) | 0.608 | 0.684 | 0.241 | 0.707 | 0.842 | 0.921 | 0.968 | 0.447 | 0.954 | 0.996 |
| (8,2) | 0.793 | 0.876 | 0.317 | 0.879 | 0.958 | 0.986 | 0.994 | 0.608 | 0.990 | 0.998 |
| (10,2) | 0.890 | 0.958 | 0.392 | 0.941 | 0.990 | 0.997 | 1.00 | 0.733 | 0.999 | 1.00 |
| (12,2) | 0.940 | 0.976 | 0.462 | 0.972 | 0.996 | 0.999 | 1.00 | 0.809 | 1.00 | 1.00 |
| | $(n_1, n_2) = (25, 50)$ | | | | | $(n_1, n_2) = (100, 100)$ | | | | |
| (4,2) | 0.335 | 0.456 | 0.166 | 0.441 | 0.636 | 0.820 | 0.854 | 0.310 | 0.853 | 0.966 |
| (6,2) | 0.648 | 0.782 | 0.285 | 0.772 | 0.932 | 0.993 | 0.998 | 0.602 | 0.997 | 1.00 |
| (8,2) | 0.826 | 0.932 | 0.394 | 0.908 | 0.986 | 0.999 | 1.00 | 0.785 | 1.00 | 1.00 |
| (10,2) | 0.912 | 0.968 | 0.474 | 0.956 | 0.996 | 1.00 | 1.00 | 0.892 | 1.00 | 1.00 |
| (12,2) | 0.953 | 0.986 | 0.553 | 0.982 | 0.996 | 1.00 | 1.00 | 0.939 | 1.00 | 1.00 |
| | $(n_1, n_2) = (50, 25)$ | | | | | $(n_1, n_2) = (200, 200)$ | | | | |
| (4,2) | 0.545 | 0.592 | 0.194 | 0.580 | 0.756 | 0.977 | 0.988 | 0.487 | 0.966 | 0.966 |
| (6,2) | 0.890 | 0.932 | 0.371 | 0.923 | 0.990 | 1.00 | 1.00 | 0.860 | 1.00 | 1.00 |
| (8,2) | 0.978 | 0.982 | 0.515 | 0.981 | 0.996 | 1.00 | 1.00 | 0.973 | 1.00 | 1.00 |
| (10,2) | 0.995 | 1.00 | 0.644 | 0.996 | 1.00 | 1.00 | 1.00 | 0.996 | 1.00 | 1.00 |
| (12,2) | 0.998 | 1.00 | 0.719 | 1.00 | 1.00 | 1.00 | 1.00 | 0.998 | 1.00 | 1.00 |

**Figure 1:** For $n_1 = 10, n_2 = 10$ powers of Z, KM, AB, PB and FP tests.

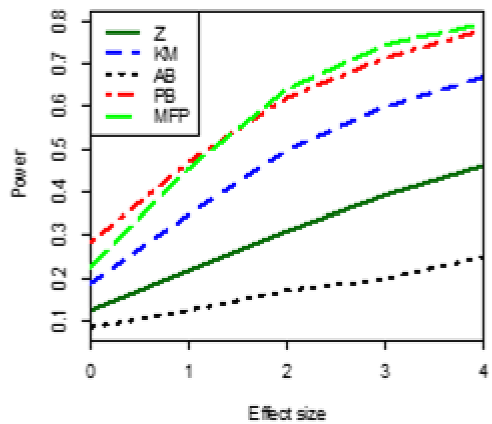



**Figure 2:** For $n_1 = 10, n_2 = 20$ powers of Z, KM, AB, PB and FP tests.

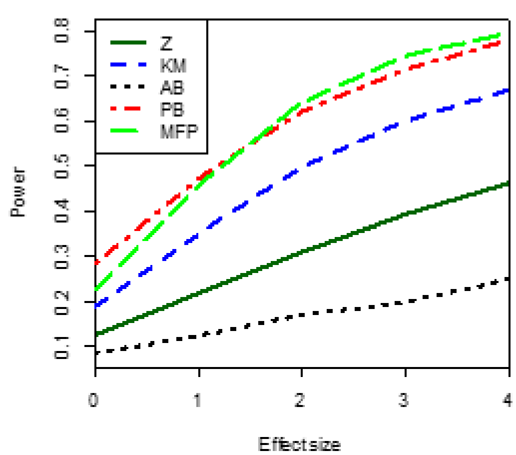


**Figure 3:** For $n_1 = 25, n_2 = 25$ powers of Z, KM, AB, PB and FP tests.

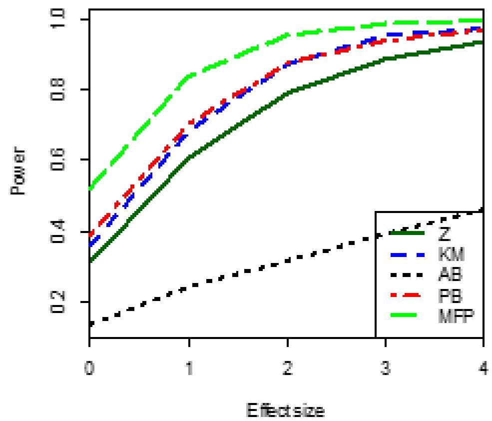



**Figure 4:** For $n_1 = 25, n_2 = 50$ powers of Z, KM, AB, PB and FP tests.

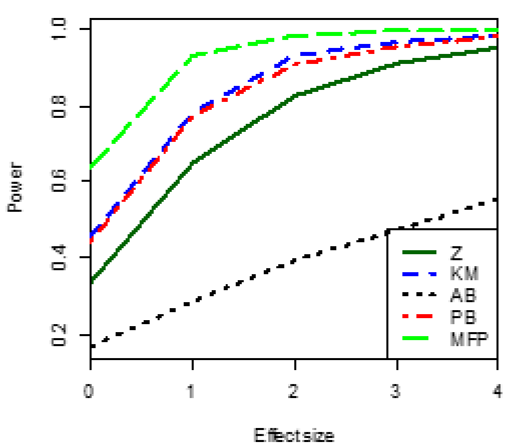


**Figure 5:** For $n_1 = 50, n_2 = 25$ powers of Z, KM, AB, PB and FP tests.

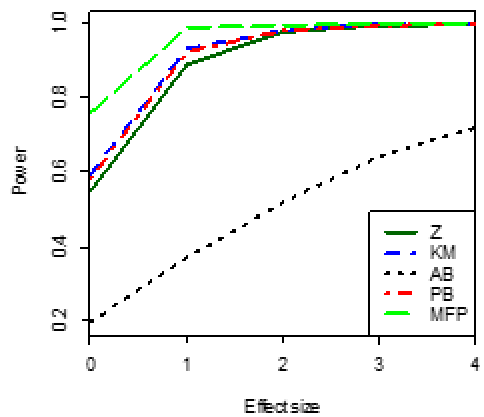
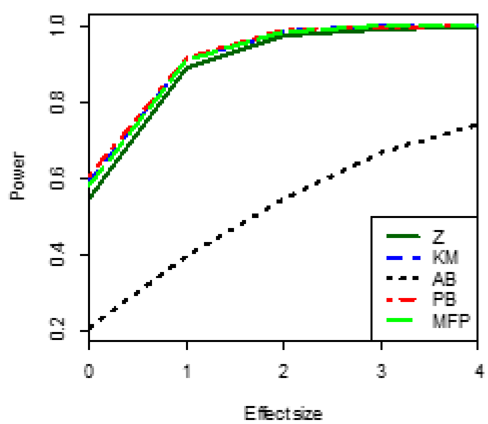



**Figure 6:** For $n_1 = 50, n_2 = 50$ powers of Z, KM, AB, PB and FP tests.




## 5. Two illustrative examples

In this section, we give two examples to illustrate the proposed approaches. The first example is about the BF problem and second example is about means of log-normal distribution.

**Example 1.** For illustrating the practical use of the proposed generalized confidence intervals, we present the result an example. The data in Table 3 is taken from Jarvis et al. [44] and Pagano and Gauvreau [45] to measure the relative level of carboxyhemoglobin for a group of nonsmokers and a group of cigarette smokers. We are interested in the interval estimation for $\mu_2 - \mu_1$ where $\mu_1$ and $\mu_2$ are the true means of carboxyhemoglobin levels for nonsmokers and cigarette smokers, respectively. The summary data is given in Table 5 and the interval limits as well as interval widths for four confidence intervals are demonstrated in Table 6. It is seen that the expected lengths of generalized approach based on algorithm 2 is shorter than the other methods.

**Table 5:** Carboxyhemoglobin for nonsmokers and smokers groups, percent

| Group | $n_i$ | $\bar{x}_i$ | $s_i^2$ |
|---|---|---|---|
| Nonsmokers | 121 | 1.3 | 1.704 |
| Smokers | 75 | 4.1 | 4.054 |

**Table 6:** % 95 confidence interval for $\mu_2 - \mu_1$

| Method | Interval limits | Interval width |
|---|---|---|
| Algorithm 1 | $(2.258, 3.330)$ | 1.072 |
| Algorithm 2 | $(2.368, 3.244)$ | 0.876 |
| GCI | $(2.283, 3.314)$ | 1.031 |

**Example 2.** We illustrate the application of the proposed tests for comparing means of two lognormal distributions. This example is discussed in Krishnamoorthy and Mathew [34] and data is the amount of rainfall (in acre-feet) from 52 clouds, of which 26 were chosen at random and seeded with silver nitrate. It was shown that normal models do not fit the data whereas lognormal models fit the data sets very well. The summary statistics for the log-transformed data are given in Table 7.

**Table 7:** The summary statistics for the log-transformed data of rainfall

| Clouds | $n_i$ | $\bar{x}_i$ | $s_i^2$ |
|---|---|---|---|
| Seeded clouds | 26 | 5.134 | 2.46 |
| Unseeded clouds | 26 | 3.990 | 2.60 |

Let $\eta_x$ and $\eta_y$ be the mean rainfall for the seeded clouds and the mean rainfall for the unseeded clouds respectively. The $p$-values for testing,

(5.1) $\quad H_0 : \eta_x = \eta_y \quad \text{vs.} \quad H_1 : \eta_x > \eta_y$

are presented in Table 8.

**Table 8:** $p$-values for testing $H_0 : \eta_x = \eta_y$ vs. $H_1 : \eta_x > \eta_y$ for the rainfall example.

| Test | $p$-values |
|---|---|
| MFB test | 0.076 |
| PB test | 0.064 |
| AB test | 0.079 |
| KM test | 0.078 |
| Z test | 0.060 |

For the significance level 5% five tests will conclude that $H_0$ is not rejected. In terms of the research question we conclude that seeded clouds do not produce more rain than unseeded clouds.

## 6. Concluding remarks

In this study, we revisited the BF problem and developed two simple methods based on generalized approach for this problem. The developed confidences intervals were compared with GCI in terms of the expected lengths and coverage probabilities. Simulation study reveal that the generalized confidence interval based on equation 2.7 has the good expected length when sample sizes are balanced or large. However, its coverage probabilities perform very poorly, especially when sample sizes are unbalance. The expected lengths of generalized confidence interval based on equation 2.6 are shorter than GCI method. Its coverage probabilities are satisfactory, especially when simple sizes are moderate and large.

We also conducted a Monte Carlo simulation study to evaluate type I error probabilities and powers of the proposed tests for comparing the means of two lognormal distributions under different scenarios. For a range of choices of the sample size and parameter configurations, we have investigated the performance of the above tests using Monte Carlo simulation. Numerical results show that when variances are unequal the size of the AB test exceeds the nominal level and its power is very weak even for large sample sizes. The type I type error rate of the Z score test underestimates the nominal level and Z test appears to be less powerful than the KM, PB and MFB tests. Regarding the problem of hypothesis testing of the ratio of two lognormal means, it is noted that the modified fiducial based test performs better than the considered tests in respect to type I error rate and power in most cases.

## References

[1] Fisher, R.A. *The fiducial argument in statistical inference*, Annals of Eugenics **6**, 391-398, 1935.

[2] Fisher, R.A. *The asymptotic approach to Behrens' integral with further tables for the d test of significance*, Annals of Eugenics **11**, 141-172, 1941.

[3] Welch, B.L. *The significance of the difference between two means when the population variances are unequal*, Biometrika **29**, 350-362, 1938.

[4] Welch, B.L. *The generalization of 'Student's' problem when several different population variances are involved*, Biometrika **34**, 28-35, 1947.

[5] Aspin, A.A. *An examination and further development of a formula arising in the problem of comparing two mean values*, Biometrika **35**, 88-96, 1948.

[6] Cochran, W.G., Cox, G.M. *Experiment Designs*, John Wiley and Sons, New York, 1950.

[7] Jing, B.Y. *Two-sample empirical likelihood method*, Statistics and Probability Letters **24**, 315-319, 1995.

[8] Kim, S.H., Cohen, A.S. *On the Behrens-Fisher problem: A review*, Journal of Educational and Behavioral Statistics **23**, 356-377, 1998.

[9] Singh, P., Saxena, K.K., Srivastava, O.P. *Power comparisons of solutions to the Behrens-Fisher problem*, American Journal of Mathematical and Management Sciences **22**, 233-250, 2002.

[10] Dong, L.B. *The Behrens-Fisher problem: an empirical likelihood approach*, Econometrics Working Paper EWP0404, University of Victoria, ISSN 1485-6441, 2004.

[11] Tsui, K. and Weerahandi, S. *Generalized p-values in significance testing of hypotheses in the presence of nuisance parameters*, Journal of the American Statistical Association **84**, 602-607, 1989.
[12] Weerahandi, S. *Generalized confidence intervals*, Journal of the American Statistical Association **88**, 899-905, 1993.
[13] Weerahandi, S. *Exact Statistical Methods for Data Analysis*, Springer, 1995.
[14] Krishnamoorthy, K., Mathew, T. *Inferences on the means of lognormal distributions using generalized p-values and generalized confidence intervals*, Journal of Statistical Planning and Inference **115**, 103-121, 2003.
[15] Krishnamoorthy, K., Lu, Y. *Inferences on the common mean of several normal populations based on the generalized variable method*, Biometrics **59**, 237-247, 2003.
[16] 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**, 5731-5742, 2007.
[17] Lee, J.C, Lin, S.H. *Generalized confidence interval for the ratio of means of two normal populations*, Journal of Statistical Planning and Inference **123**, 49-60, 2004.
[18] Chang, C.H., Pal, N. *A revisit to the Behrens-Fisher problem: comparison of five test methods*, Communications in Statistics-Simulation and Computation **37**, 1064-1085, 2008.
[19] Zheng, S., Shi, N.Z., Ma, W. *Statistical inference on difference or ratio of means from heteroscedastic normal populations*, Journal of Statistical Planning and Inference **140**, 1236-1242, 2010.
[20] Ozkip, E., Yazici, B., Sezer, A. *A simulation study on tests for the Behrens-Fisher problem*, Turkiye Klinikleri: Journal of Biostatistics **6**, 59-66, 2014.
[21] Sezer, A., Ozkip, E., Yazici, B. *Comparison of confidence intervals for the Behrens-Fisher problem*, Communications in Statistics-Simulation and Computation, accepted manuscript, 2015.
[22] Ye, R.D., Ma, T.F., Luo, K. *Inferences on the reliability in balanced and unbalanced one-way random models*, Journal of Statistical Computation and Simulation **84**, 1136-1153, 2014.
[23] Gunasekera, S., Ananda, M. *Generalized variable method inference for the location parameter of the general half-normal distribution*, Journal of Statistical Computation and Simulation **85**, 2115-2132, 2015.
[24] Zhao, G., Xu, X. *The p-values for one-sided hypothesis testing in univariate linear calibrations*, Communications in Statistics - Simulation and Computation, accepted manuscript, 2016.
[25] Krishnamoorthy, K., Mathew, T., Ramachandran, G. *Generalized p-values and confidence intervals: A novel approach for analyzing lognormally distributed exposure data*, , Journal of Occupational and Environmental Hygiene **3**, 642-650, 2006.
[26] Shen, H., Brown, L., Hui, Z. *Efficient estimation of log-normal means with application to pharmacokinetics data*, Statistics in Medicine **25**, 3023-3038, 2006.
[27] Land, C.E. *Confidence intervals for linear functions of the lognormal mean and variance*, Annals of Mathematical Statistics **42**, 1187-1205, 1971.
[28] Land, C.E. *An evaluation of approximate confidence interval estimation methods for lognormal means*, Technometrics **14**, 145-158, 1972.
[29] Land, C.E. *Standard confidence limits for linear functions of the normal mean and variance*, Journal of the American Statistical Association **68**, 960-963, 1973.
[30] Angus ,J.E. *Inferences on the lognormal mean for complete samples*, Comm. Statist. Simulation Comput. **17**, 1307-1331, 1988.
[31] Angus ,J.E. *Bootstrap one-sided confidence intervals for the lognormal mean*, Statistician **43**, 395-401, 1994.
[32] Zhou, X.H., Gao, S., Hui, S.L. *Methods for comparing the means of two independent lognormal samples*, Biometrics **53**, 1129-1135, 1997.
[33] Wu, J., Jiang, G., Wong, A.C.M., Sun, X. *Likelihood analysis for the ratio of means of two independent log-normal distributions*, Biometrics **58**, 463-469, 2002.
[34] Krishnamoorthy, K., Mathew, T. *Inferences on the means of lognormal distributions using generalized p-values and generalized confidence intervals*, Journal of Statistical Planning and Inference **115**, 130-121, 2003.

[35] Chen, Y.H., Zhou, X.H. *Interval estimates for the ratio and difference of two lognormal means*, Statistics in Medicine **25**, 4099-4113, 2006.

[36] Hannig, J., Iyer, H., Patterson, P. *Fiducial generalized confidence intervalss*, Journal of the American Statistical Association **101**, 254-269, 2006.

[37] Li, X., Wang, J., Liang, H. *Comparison of several means: A fiducial based approach*, Computational Statistics and Data Analysis **55**, 1993-2002, 2011.

[38] Abdollahnezhad, K., Babanezhad, M., Jafari, A.A. *Inference on difference of means of two log-normal distributions a generalized approach*, Journal of Statistical and Econometric Methods **1**, 125-131, 2012.

[39] Weng, S.F., Myers, L. *Performance of confidence interval tests for the ratio of two lognormal means applied to Weibull and gamma distribution*, Journal of Statistical Computation and Simulation **82**, 1035-1054, 2012.

[40] Jafari, A.A., Abdollahnezhad, K. *Inferences on the means of two log-normal distributions; A computational approach test*, Communications in Statistics-Simulation and Computation **44**, 1659-1672, 2014.

[41] Jiang, L., Rekkas, M., Wong, A. *Comparing the means of two log-normal distributions: A likelihood approach*, Journal of Statisical and Econemtric Methods **3**, 137-152, 2014.

[42] Krishnamoorthy, K., Lu, F., Mathew, T. *A parametric bootstrap approach for ANOVA with unequal variances: fixed and random models*, Comput. Stat. Data Anal. **51**, 5731-5742, 2007.

[43] Fisher, R.A. *The fiducial argument in statistical inference*, Annals of Eugenics **6**, 391-398, 1935.

[44] Jarvis, M.J., Tunstall-Pedoe, H., Feyerabend, C., Vesey, C., Saloojee, Y. *Comparison of tests used to distinguish smokers from nonsmokers*, American Journal of Public Health **77**, 1435-1438, 1987.

[45] Pagano, M., Gauvreau, K. *Principles of Biostatistics*, California: Cuxbury, 1993.