Analysis of Medical Data Using Interval Estimators for Common Mean of Gaussian Distributions with Unknown Coefficients of Variation

Warisa Thangjai[1], Sa-Aat Niwitpong[2], and Suparat Niwitpong[2](✉)

[1] Department of Statistics, Faculty of Science, Ramkhamhaeng University, Bangkok 10240, Thailand

[2] Department of Applied Statistics, Faculty of Applied Science, King Mongkut's University of Technology North Bangkok, Bangkok 10800, Thailand

{sa-aat.n,suparat.n}@sci.kmutnb.ac.th

Abstract. The common mean of Gaussian distributions is a parameter of interest when analyzing medical data. In practice, the population coefficient of variation (CV) is unknown because the population mean and variance are unknown. In this study, the common mean of Gaussian distributions with unknown CVs is considered and four new interval estimators for it using generalized confidence interval (GCI), large sample (LS), adjusted method of variance estimates recovery (adjusted MOVER), and standard bootstrap (SB) approaches are proposed. Furthermore, the proposed interval estimators are compared with a previously reported one based on the GCI approach. Monte Carlo simulation was used to evaluate the performances of the interval estimators based on their coverage probabilities and average lengths, while, medical datasets were used to illustrate the efficacy of these approaches. Our findings show that the interval estimator based on the GCI approach for the common mean of Gaussian distributions with unknown CVs provided the best performance in terms of coverage probability for all sample sizes. However, the adjusted MOVER and SB approaches can be considered as an alternative when the sample size is large ($n_i \geq 100$).

Keywords: Adjusted MOVER approach · CV · GCI approach · Mean · SB approach

1 Introduction

The population coefficient of variation (CV), which is free from a unit of measurement, is defined as the ratio of the population standard deviation to the population mean, $\tau = \sigma/\mu$, and has been widely applied in many fields, e.g., agriculture, biology, and environmental and physical sciences. Estimating a known CV has been suggested by many scholars. For example, Gerig and Sen [1] used Canadian migratory bird survey data from 1969 and 1970 while assuming that the CV for

K. Honda et al. (Eds.): IUKM 2022, LNAI 13199, pp. 155–166, 2022.
https://doi.org/10.1007/978-3-030-98018-4_13

each province was known. Meanwhile, estimating the mean of Gaussian distributions with a known CV has also been studied extensively (e.g., Searls [2] and Niwitpong [3]). However, the CV is unknown when the population mean and variance have been estimated and thus, needs to be estimated too. Srivastava [4] and Srivastava [5] proposed an estimator for the normal population mean with an unknown CV and indicated that it is more efficient than a previously reported sample mean estimator. He later presented a uniformly minimum variance unbiased estimator of the efficiency ratio and compared its usefulness to estimate an unknown CV with an existing estimator (Srivastava and Singh [6]). Sahai [7] provided an estimator for the normal mean with unknown CV and studied it along the same lines as Srivastava [4] and Srivastava [5] estimators. Meanwhile, Thangjai et al. [8] presented confidence intervals for the normal mean and the difference between two normal means with unknown CVs. In addition, Thangjai et al. [9] proposed the Bayesian confidence intervals for means of normal distributions with unknown CVs.

In practice, samples are collected at different time points, and the problem of estimating common parameters under these circumstances has been widely studied by several researchers. Krishnamoorthy and Lu [10] proposed the generalized variable approach for inference on the common mean of normal distributions. Lin and Lee [11] developed a new generalized pivotal quantity based on the best linear unbiased estimator for constructing confidence intervals for the common mean of normal distributions. Tian [12] presented procedures for inference on the common CV of normal distributions. Tian and Wu [13] provided the generalized variable approach for inference on the common mean of log-normal distributions. Thangjai et al. [14] investigated a new confidence interval for the common mean of normal distributions using the adjusted method of variance estimates recovery (adjusted MOVER) approach. Finally, the estimator of Srivastava [4] is well established for constructing confidence intervals for the common mean of Gaussian distributions with unknown CVs.

Interval estimators for the common mean of Gaussian distributions with unknown CVs have been proposed in several medical science studies, such as the common percentage of albumin in human plasma proteins from four sources (Jordan and Krishnamoorthy [15]) and quality assurance in medical laboratories for the diagnostic determinations of hemoglobin, red blood cells, the mean corpuscular volume, hematocrit, white blood cells, and platelets in normal and abnormal blood samples (Tian [12] and Fung and Tsang [16]).

Herein, the concepts in Thangjai et al. [8] and Thangjai and Niwitpong [17] are extended to k populations to construct new interval estimators for the common mean of Gaussian distributions with unknown CVs. The approaches to construct these interval estimators: the generalized confidence interval (GCI), large sample (LS), adjusted MOVER, and standard bootstrap (SB) are compared with the GCI approach of Lin and Lee [11]. The GCI approach first introduced by Weerahandi [18] has been used successfully to construct interval estimators (e.g., Krishnamoorthy and Lu [10], Lin and Lee [11], Tian [12], Tian and Wu [13], Ye et al. [19]). The LS approach using the central limit theorem (along with a GCI approach) was first proposed by Tian and Wu [13] to construct confidence intervals for the common mean of log-normal distributions. The adjusted MOVER approach motivated by Zou and Donner [20] and Zou et al. [21] was extended by Thangjai et al. [14] and Thangjai and Niwitpong [17] to construct an interval estimator for a common parameter.

2 Preliminaries

In this section, the lemma and theorem are explained to estimate the interval estimators for common Gaussian mean with unknown CVs.

Let $X = (X_1, X_2, ..., X_n)$ be a random variable from the Gaussian distribution with mean μ and variance σ^2. The population CV is $\tau = \sigma/\mu$. Let $\bar{X}$ and S^2 be sample mean and sample variance for X, respectively. The CV estimator is $\hat{\tau} = S/\bar{X}$.

Following Srivastava [4] and Thangjai et al. [8], the Gaussian mean estimator when the CV is unknown, $\hat{\theta}$, is

$$\hat{\theta} = \frac{n\bar{X}}{n + \frac{S^2}{\bar{X}^2}}. \tag{1}$$

According to Thangjai et al. [8], the mean and variance of the mean estimator with unknown CV are

$$E\left(\hat{\theta}\right) = \left(\frac{\mu}{1 + \left(\frac{\sigma^2}{n\mu^2+\sigma^2}\right)\left(1 + \frac{2\sigma^4+4n\mu^2\sigma^2}{(n\mu^2+\sigma^2)^2}\right)}\right) * \left(1 + \frac{\left(\frac{n\sigma^2}{n\mu^2+\sigma^2}\right)^2\left(\frac{2}{n} + \frac{2\sigma^4+4n\mu^2\sigma^2}{(n\mu^2+\sigma^2)^2}\right)}{\left(n + \left(\frac{n\sigma^2}{n\mu^2+\sigma^2}\right)\left(1 + \frac{2\sigma^4+4n\mu^2\sigma^2}{(n\mu^2+\sigma^2)^2}\right)\right)^2}\right) \tag{2}$$

and

$$Var\left(\hat{\theta}\right) = \left(\frac{\mu}{1 + \left(\frac{\sigma^2}{n\mu^2+\sigma^2}\right)\left(1 + \frac{2\sigma^4+4n\mu^2\sigma^2}{(n\mu^2+\sigma^2)^2}\right)}\right)^2 * \left(\frac{\sigma^2}{n\mu^2} + \frac{\left(\frac{n\sigma^2}{n\mu^2+\sigma^2}\right)^2\left(\frac{2}{n} + \frac{2\sigma^4+4n\mu^2\sigma^2}{(n\mu^2+\sigma^2)^2}\right)}{\left(n + \left(\frac{n\sigma^2}{n\mu^2+\sigma^2}\right)\left(1 + \frac{2\sigma^4+4n\mu^2\sigma^2}{(n\mu^2+\sigma^2)^2}\right)\right)^2}\right). \tag{3}$$

Consider k independent Gaussian distributions with a common mean with unknown CVs. Let $X_i = (X_{i1}, X_{i2}, ..., X_{in_i})$ be a random variable from the i-th Gaussian distribution with the common mean μ and possibly unequal variances σ_i^2 as follows: $X_{ij} \sim N(\mu, \sigma_i^2)$; $i = 1, 2, ..., k$, $j = 1, 2, ..., n_i$.

For the i-th sample, let $\bar{X}_i$ and $\bar{x}_i$ be sample mean and observed sample mean of X_i, respectively. And let S_i^2 and s_i^2 be sample variance and observed sample variance of X_i, respectively. According to Thangjai et al. [8], the estimator of Srivastava [4] is well established. The estimator is given by

$$\hat{\theta}_i = \frac{n_i \bar{X}_i}{n_i + \frac{S_i^2}{(\bar{X}_i)^2}}; i = 1, 2, ..., k. \tag{4}$$

This paper is interested in constructing confidence intervals for the common Gaussian mean with unknown CVs, based on Graybill and Deal [22], defined as follows:

$$\hat{\theta} = \sum_{i=1}^{k} \frac{\hat{\theta}_i}{\widetilde{Var}\left(\hat{\theta}_i\right)} \Bigg/ \sum_{i=1}^{k} \frac{1}{\widetilde{Var}\left(\hat{\theta}_i\right)}, \tag{5}$$

where $\widetilde{Var}\left(\hat{\theta}_i\right)$ denotes the estimator of $Var\left(\hat{\theta}_i\right)$ which is defined in Eq. (3) with μ_i and σ_i^2 replaced by $\bar{x}_i$ and s_i^2, respectively.

2.1 GCI

Definition 1. Let $X = (X_1, X_2, ..., X_n)$ be a random variable from a distribution $F(x|\delta)$, where $x = (x_1, x_2, ..., x_n)$ be an observed sample, $\delta = (\theta, \nu)$ is a unknown parameter vector, θ is a parameter of interest, and ν is a nuisance parameters. Let $R = R(X; x, \delta)$ be a function of X, x and δ. The random quantity R is called a generalized pivotal quantity if it satisfies the following two properties; see Weerahandi [18]:

(i) The probability distribution of R is free of unknown parameters.
(ii) The observed value of R does not depend on the vector of nuisance parameters.

The $100(\alpha/2)$-th and $100(1-\alpha/2)$-th percentiles of R are the lower and upper limits of $100(1-\alpha)\%$ two-sided GCI.

Following Thangjai et al. [8], the generalized pivotal quantities of σ_i^2, μ_i, and θ_i based on the i-th sample are defined as follows:

$$R_{\sigma_i^2} = \frac{(n_i - 1)\, s_i^2}{V_i}. \tag{6}$$

$$R_{\mu_i} = \bar{x}_i - \frac{Z_i}{\sqrt{U_i}} \sqrt{\frac{(n_i - 1)\, s_i^2}{n_i}} \tag{7}$$

and

$$R_{\theta_i} = \frac{n_i R_{\mu_i}}{n_i + \frac{R_{\sigma_i^2}}{(R_{\mu_i})^2}}, \tag{8}$$

where V_i denotes a chi-squared distribution with $n_i - 1$ degrees of freedom, Z_i denotes a standard normal distribution, and U_i denotes a chi-squared distribution with $n_i - 1$ degrees of freedom.

According to Tian and Wu [13], the generalized pivotal quantity for the common Gaussian mean with unknown CVs is a weighted average of the generalized pivotal quantity. That is given by

$$R_\theta = \sum_{i=1}^{k} \frac{R_{\theta_i}}{R_{Var(\hat{\theta}_i)}} \Bigg/ \sum_{i=1}^{k} \frac{1}{R_{Var(\hat{\theta}_i)}}, \tag{9}$$

where $R_{Var(\hat{\theta}_i)}$ is defined in Eq. (3) with μ_i and σ_i^2 replaced by R_{μ_i} and $R_{\sigma_i^2}$, respectively.

Hence, the R_θ is the generalized pivotal quantity for θ and is satisfied the conditions (i) and (ii) in Definition 1. Then the common Gaussian mean with unknown CVs can be constructed from R_θ.

Therefore, the $100(1-\alpha)\%$ two-sided confidence interval for the common Gaussian mean with unknown CVs based on the GCI approach is

$$CI_{GCI} = [L_{GCI}, U_{GCI}] = [R_\theta\left(\alpha/2\right), R_\theta\left(1-\alpha/2\right)], \tag{10}$$

where $R_\theta\left(\alpha/2\right)$ and $R_\theta\left(1-\alpha/2\right)$ denote the $100(\alpha/2)$-th and $100(1-\alpha/2)$-th percentiles of R_θ, respectively.

2.2 LS Confidence Interval

According to Graybill and Deal [22] and Tian and Wu [13], the LS estimate of the Gaussian mean with unknown CV is a pooled estimated estimator of the Gaussian mean with unknown CV defined as in Eq. (5), where $\hat{\theta}_i$ is defined in Eq. (4) and $\widetilde{Var}\left(\hat{\theta}_i\right)$ denotes the estimator of $Var\left(\hat{\theta}_i\right)$ which is defined in Eq. (3) with μ_i and σ_i^2 replaced by $\bar{x}_i$ and s_i^2, respectively.

The distribution of $\hat{\theta}$ is approximately Gaussian distribution when the sample size is large. Then the quantile of the Gaussian distribution is used to construct confidence interval for θ. Therefore, the $100(1-\alpha)\%$ two-sided confidence interval for the common Gaussian mean with unknown CVs based on the LS approach is

$$\begin{aligned} CI_{LS} &= [L_{LS}, U_{LS}] \\ &= [\hat{\theta} - z_{1-\alpha/2}\sqrt{1 \Bigg/ \sum_{i=1}^{k} \frac{1}{\widetilde{Var}\left(\hat{\theta}_i\right)}}, \hat{\theta} + z_{1-\alpha/2}\sqrt{1 \Bigg/ \sum_{i=1}^{k} \frac{1}{\widetilde{Var}\left(\hat{\theta}_i\right)}}], \end{aligned} \tag{11}$$

where $z_{1-\alpha/2}$ denotes the $(1-\alpha/2)$-th quantile of the standard normal distribution.

2.3 Adjusted MOVER Confidence Interval

Now recall that Z is a standard normal distribution with the mean 0 and variance 1, defined as follows:

$$Z = \frac{\bar{X} - \mu}{\sqrt{\widetilde{Var}\left(\hat{\theta}\right)}} \sim N(0,1). \tag{12}$$

The confidence interval for mean of Gaussian distribution is

$$CI_{\mu} = [l,u] = [\bar{x} - z_{1-\alpha/2}\sqrt{\widetilde{Var}\left(\hat{\theta}\right)}, \bar{x} + z_{1-\alpha/2}\sqrt{\widetilde{Var}\left(\hat{\theta}\right)}]. \tag{13}$$

For $i = 1, 2, ..., k$, the lower limit (l_i) and upper limit (u_i) for the normal mean μ_i based on the i-th sample can be defined as

$$l_i = \bar{x}_i - z_{1-\alpha/2}\sqrt{\widetilde{Var}\left(\hat{\theta}_i\right)} \tag{14}$$

and

$$u_i = \bar{x}_i + z_{1-\alpha/2}\sqrt{\widetilde{Var}\left(\hat{\theta}_i\right)}, \tag{15}$$

where $\widetilde{Var}\left(\hat{\theta}_i\right)$ denotes the estimator of $Var\left(\hat{\theta}_i\right)$ which is defined in Eq. (3) and $z_{1-\alpha/2}$ denotes the $(1-\alpha/2)$-th quantile of the standard normal distribution.

According to Thangjai et al. [14] and Thangjai and Niwitpong [17], the common mean with unknown CVs is weighted average of the mean with unknown CV $\hat{\theta}_i$ based on k individual samples. The common mean with unknown CVs has the following form

$$\hat{\theta} = \sum_{i=1}^{k} \frac{\hat{\theta}_i}{\widehat{Var}\left(\hat{\theta}_i\right)} \Bigg/ \sum_{i=1}^{k} \frac{1}{\widehat{Var}\left(\hat{\theta}_i\right)}, \tag{16}$$

where $\hat{\theta}_i$ is defined in Eq. (4), $\widehat{Var}\left(\hat{\theta}_i\right) = \frac{1}{2}(\frac{(\hat{\theta}_i - l_i)^2}{z_{\alpha/2}^2} + \frac{(u_i - \hat{\theta}_i)^2}{z_{\alpha/2}^2})$, and l_i and u_i are defined in Eqs. (14) and (15), respectively.

Therefore, the $100(1-\alpha)\%$ two-sided confidence interval for the common Gaussian mean with unknown CVs based on the adjusted MOVER approach is

$$\begin{aligned} CI_{AM} &= [L_{AM}, U_{AM}] \\ &= [\hat{\theta} - z_{1-\alpha/2}\sqrt{1 \Bigg/ \sum_{i=1}^{k} \frac{z_{\alpha/2}^2}{(\hat{\theta}_i - l_i)^2}}, \hat{\theta} + z_{1-\alpha/2}\sqrt{1 \Bigg/ \sum_{i=1}^{k} \frac{z_{\alpha/2}^2}{(u_i - \hat{\theta}_i)^2}}], \end{aligned} \tag{17}$$

where $\hat{\theta}$ is defined in Eq. (16), and $z_{\alpha/2}$ and $z_{1-\alpha/2}$ denote the $(\alpha/2)$-th and $(1-\alpha/2)$-th quantiles of the standard normal distribution, respectively.

2.4 SB Confidence Interval

Let $X_i^* = (X_{i1}^*, X_{i2}^*, ..., X_{in_i}^*)$ be a bootstrap sample with replacement from $X_i = (X_{i1}, X_{i2}, ..., X_{in_i})$ and let $\bar{X}_i^*$ and S_i^{2*} be mean and variance of X_i^*, respectively. Let $x_i^* = (x_{i1}^*, x_{i2}^*, ..., x_{in_i}^*)$ be an observed value of $X_i^* = (X_{i1}^*, X_{i2}^*, ..., X_{in_i}^*)$ and let $\bar{x}_i^*$ and s_i^{2*} be mean and variance of x_i^*, respectively. The estimates of $\hat{\theta}_i^*$ and $Var(\hat{\theta}_i^*)$ are

$$\hat{\theta}_i^* = \frac{n_i \bar{X}_i^*}{n_i + \frac{S_i^{2*}}{(\bar{X}_i^*)^2}} \tag{18}$$

and

$$Var\left(\hat{\theta}_i^*\right) = \left(\frac{\mu_i^*}{1 + \left(\frac{\sigma_i^{2*}}{n_i(\mu_i^*)^2 + \sigma_i^{2*}}\right)\left(1 + \frac{2\sigma_i^{4*} + 4n_i(\mu_i^*)^2\sigma_i^{2*}}{\left(n_i(\mu_i^*)^2 + \sigma_i^{2*}\right)^2}\right)}\right)^2 * \left(\frac{\sigma_i^{2*}}{n_i(\mu_i^*)^2} + \frac{\left(\frac{n_i\sigma_i^{2*}}{n_i(\mu_i^*)^2 + \sigma_i^{2*}}\right)^2\left(\frac{2}{n_i} + \frac{2\sigma_i^{4*} + 4n_i(\mu_i^*)^2\sigma_i^{2*}}{\left(n_i(\mu_i^*)^2 + \sigma_i^{2*}\right)^2}\right)}{\left(n_i + \left(\frac{n_i\sigma_i^{2*}}{n_i(\mu_i^*)^2 + \sigma_i^{2*}}\right)\left(1 + \frac{2\sigma_i^{4*} + 4n_i(\mu_i^*)^2\sigma_i^{2*}}{\left(n_i(\mu_i^*)^2 + \sigma_i^{2*}\right)^2}\right)\right)^2}\right). \tag{19}$$

According to Graybill and Deal [22], the common Gaussian mean with unknown CVs is a pooled estimated unbiased estimator of the Gaussian mean with unknown CVs based on k individual samples. The common Gaussian mean with unknown CVs is defined by

$$\hat{\theta}^* = \sum_{i=1}^{k} \frac{\hat{\theta}_i^*}{\widetilde{Var}\left(\hat{\theta}_i^*\right)} \Bigg/ \sum_{i=1}^{k} \frac{1}{\widetilde{Var}\left(\hat{\theta}_i^*\right)}, \tag{20}$$

where $\hat{\theta}_i^*$ is defined in Eq. (18) and $\widetilde{Var}\left(\hat{\theta}_i^*\right)$ is the estimator of $Var\left(\hat{\theta}_i^*\right)$ which is defined in Eq. (19) with μ_i^* and σ_i^{2*} replaced by $\bar{x}_i^*$ and s_i^{2*}, respectively.

The B bootstrap statistics are used to construct the sampling distribution for estimating the confidence interval for the common Gaussian mean with unknown CVs. Therefore, the $100(1-\alpha)\%$ two-sided confidence interval for the common Gaussian mean with unknown CVs based on the SB approach is

$$CI_{SB} = [L_{SB}, U_{SB}] = [\bar{\hat{\theta}}^* - z_{1-\alpha/2} S_{\hat{\theta}^*}, \bar{\hat{\theta}}^* + z_{1-\alpha/2} S_{\hat{\theta}^*}], \tag{21}$$

where $\bar{\hat{\theta}}^*$ and $S_{\hat{\theta}^*}$ are the mean and standard deviation of $\hat{\theta}^*$ defined in Eq. (20) and $z_{1-\alpha/2}$ denotes the $100(1-\alpha/2)$-th percentile of the standard normal distribution.

Next, we briefly review the GCI of Lin and Lee [11] for the common mean of Gaussian distributions. The generalized pivotal quantity based on the best linear un-biased estimator for the common Gaussian mean μ is

$$R_\mu = \frac{\sum_{i=1}^{k} \frac{n_i \bar{x}_i U_i}{v_i} - Z\sqrt{\sum_{i=1}^{k} \frac{n_i U_i}{v_i}}}{\sum_{i=1}^{k} \frac{n_i U_i}{v_i}}, \quad (22)$$

where Z denotes the standard normal distribution, U_i denotes a chi-squared distribution with $n_i - 1$ degrees of freedom, and $v_i = (n_i - 1)s_i^2$.

Therefore, the $100(1-\alpha)\%$ two-sided confidence interval for the common Gaussian mean based on the GCI approach of Lin and Lee [20] is

$$C_{LL} = [L_{LL}, U_{LL}] = [R_\mu(\alpha/2), R_\mu(1-\alpha/2)], \quad (23)$$

where $R_\mu(\alpha/2)$ and $R_\mu(1-\alpha/2)$ denote the $100(\alpha/2)$-th and $100(1-\alpha/2)$-th percentiles of R_μ, respectively.

3 Simulation Studies

Monte Carlo simulation was used to estimate the coverage probabilities (CPs) and the average lengths (ALs) of all confidence intervals; those constructed via the GCI, LS, adjusted MOVER, and SB approaches are denoted as CI_{GCI}, CI_{LS}, CI_{AM}, and CI_{SB}, respectively, while the GCI of Lin and Lee [20] is denoted as CI_{LL}. The CP of the $100(1-\alpha)\%$ confidence level is $c \pm z_{\alpha/2}\sqrt{\frac{c(1-c)}{M}}$, where c is the nominal confidence level and M is the number of simulation runs. At the 95% confidence level, the best performing confidence interval will have a CP in the range [0.9440,0.9560] with the shortest AL.

Each confidence interval was evaluated at the nominal confidence level of 0.95. The number of populations $k = 2$; and the sample sizes within each population n_1 and n_2 were given in the following table. Without loss of generality (Thangjai et al. [8]), the common mean of Gaussian data within each population was $\mu = 1.0$. The population standard deviations were set at $\sigma_1 = 0.5$, 1.0, 1.5, 2.0 and $\sigma_2 = 1.0$. The CVs were computed by $\tau_i = \sigma_i/\mu$, where $i = 1, 2$. Hence, the ratio of τ_1 to τ_2 was reduced to σ_1/σ_2.

The result of simulations with the number of simulation runs $M = 5{,}000$ is reported in Table 1. Only CI_{GCI} obtained CPs greater than 0.95 in all cases whereas those of CI_{LS}, CI_{AM}, CI_{SB}, and CI_{LL} were under 0.95. However, the CPs of CI_{LS}, CI_{AM}, and CI_{SB} increased and became close to 0.95 when the sample size was increased. For $n_i \leq 50$, the CPs of CI_{LS}, CI_{AM}, and CI_{SB} tended to decrease when σ_1/σ_2 increased. Moreover, the CPs of CI_{GCI} did not change when σ_1/σ_2 was varied. Hence, CI_{GCI} is preferable for most cases, while CI_{AM} and CI_{SB}, which are easy to use in practice, can be used when the sample size is large ($n_i \geq 100$).

As the sample case (k) increased, CI_{GCI} is preferable when the sample size is small. For a large sample size, CI_{GCI}, CI_{AM}, CI_{SB}, and CI_{LL} performed similarly in terms of CP but the ALs of the CI_{AM}, CI_{SB}, and CI_{LL} were shorter than that of CI_{GCI}.

Table 1. The CPs and ALs of 95% two-sided confidence intervals for the common mean of Gaussian distributions with unknown CVs: 2 sample cases.

n_1	n_2	μ	σ_1/σ_2	CP (AL)				
				CI_{GCI}	CI_{LS}	CI_{AM}	CI_{SB}	CI_{LL}
30	30	1.0	0.5	**0.9532**	0.9338	0.9432	0.9412	0.9450
				(0.3586)	(0.3176)	(0.3313)	(0.3299)	(0.3277)
			1.0	**0.9608**	0.9350	**0.9506**	0.9476	0.9460
				(0.6305)	(0.5297)	(0.5526)	(0.5498)	(0.5136)
			1.5	**0.9580**	0.9232	0.9456	**0.9526**	0.9492
				(0.7689)	(0.6473)	(0.6744)	(0.7130)	(0.6065)
			2.0	**0.9574**	0.8822	0.9140	0.9480	0.9474
				(0.8126)	(0.6707)	(0.6953)	(0.8183)	(0.6563)
50	50	1.0	0.5	0.9496	0.9388	0.9478	0.9418	0.9460
				(0.2706)	(0.2474)	(0.2536)	(0.2501)	(0.2515)
			1.0	**0.9642**	0.9480	**0.9562**	0.9474	0.9480
				(0.4506)	(0.4075)	(0.4177)	(0.4061)	(0.3962)
			1.5	**0.9524**	0.9426	**0.9544**	0.9496	0.9430
				(0.5505)	(0.5020)	(0.5145)	(0.5110)	(0.4658)
			2.0	**0.9508**	0.9116	0.9340	**0.9506**	0.9428
				(0.6042)	(0.5269)	(0.5388)	(0.6109)	(0.5024)
100	100	1.0	0.5	**0.9540**	0.9478	**0.9512**	0.9490	**0.9504**
				(0.1882)	(0.1751)	(0.1772)	(0.1755)	(0.1764)
			1.0	**0.9588**	**0.9520**	**0.9554**	0.9498	0.9492
				(0.2980)	(0.2832)	(0.2867)	(0.2806)	(0.2786)
			1.5	**0.9506**	**0.9570**	**0.9586**	0.9496	0.9494
				(0.3623)	(0.3494)	(0.3536)	(0.3363)	(0.3279)
			2.0	**0.9568**	**0.9550**	**0.9622**	**0.9580**	**0.9506**
				(0.3892)	(0.3748)	(0.3793)	(0.3831)	(0.3536)
200	200	1.0	0.5	**0.9500**	0.9476	**0.9500**	0.9476	0.9482
				(0.1322)	(0.1239)	(0.1247)	(0.1240)	(0.1244)
			1.0	0.9498	0.9478	0.9498	0.9462	0.9474
				(0.2030)	(0.1984)	(0.1997)	(0.1972)	(0.1968)
			1.5	0.9488	0.9490	**0.9514**	0.9418	0.9428
				(0.2543)	(0.2415)	(0.2430)	(0.2334)	(0.2310)
			2.0	**0.9592**	**0.9616**	**0.9638**	**0.9542**	**0.9566**
				(0.2649)	(0.2625)	(0.2641)	(0.2547)	(0.2485)

4 Empirical Application

Empirical application of the proposed confidence intervals to real data were presented and compared with CI_{LL}.

The dataset reported by Fung and Tsang [16] and Tian [12] and used here comprises hemoglobin, red blood cells, the mean corpuscular volume, hematocrit, white blood cells, and platelet values in normal and abnormal blood samples collected by the Hong Kong Medical Technology Association in 1995 and 1996. The summary statistics for 1995 are $\bar{x}_1 = 84.1300$, $s_1^2 = 3.3900$, and $n_1 = 63$, and those for 1996 are $\bar{x}_2 = 85.6800$, $s_2^2 = 2.9460$, and $n_2 = 72$. The means of the Gaussian distributions with unknown CVs are $\hat{\theta}_1 = 84.1294$ and $\hat{\theta}_2 = 85.6795$ for 1995 and 1996, respectively, while the common mean of the Gaussian distributions with unknown CVs is $\hat{\theta} = 85.1962$.

The two datasets fit Gaussian distributions. The 95% two-sided confidence intervals for CI_{GCI}, CI_{LS}, CI_{AM}, and CI_{SB} were [84.1099,85.8884], [61.4262,108.9661], [60.9972,109.3992], and [84.4604,85.6013] with interval lengths of 1.7785, 47.5399, 48.4020, and 1.1409. For comparison, CI_{LL} provided [84.6502,85.3635] with an interval length of 0.7133. Thus, CI_{LL} had the shortest interval length, while CI_{SB} performed the best out of the proposed approaches as its interval length was shorter than those of the other three for $k = 2$.

Therefore, these results confirm our simulation study in the previous section in term of length. In simulation, the GCI of Lin and Lee [20] is the shortest average lengths, but the coverage probabilities are less than the nominal confidence level of 0.95. Furthermore, the coverage probability and length in this example are computed by using only one sample, whereas the coverage probability and average length in the simulation are computed by using 5,000 random samples. Therefore, the GCI of Lin and Lee [20] is not recommended to construct the confidence intervals for common mean of Gaussian distributions with unknown CVs.

5 Discussion and Conclusions

Thangjai et al. [8] proposed confidence intervals for the mean and difference of means of normal distributions with unknown coefficients of variation. In addition, Thangjai et al. [9] presented the Bayesian approach to construct the confidence intervals for means of normal distributions with unknown coefficients of variation. In this paper, we extend the work of Thangjai et al. [8,9] to construct confidence intervals for the common mean of k Gaussian distributions with unknown CVs.

Herein, GCI, LS, adjusted MOVER, and SB approaches to construct interval estimators for the common mean of Gaussian distributions with unknown CVs are presented. Their CPs and ALs were evaluated via a Monte Carlo simulation and compared with the confidence interval based on the GCI approach of Lin and Lee [11]. The results of the simulation studies indicate that the confidence intervals performed similarly based on their CPs for large sample sizes (i.e.,

$n_i \geq 100$). However, the CP of CI_{GCI} was more satisfactory than those of the other confidence intervals. Moreover, the CPs of CI_{AM}, CI_{SB}, and CI_{LL} were close to 0.95 and their ALs were slightly shorter than CI_{GCI} when the sample size was large (i.e., $n_i \geq 100$). Thus, CI_{AM} and CI_{SB} can be considered as an alternative to construct an interval estimator for the common mean of Gaussian distributions with unknown CVs when the sample size is large whereas CI_{LS} is not recommended for small sample sizes (i.e., $n_i < 100$) as its CP is below 0.95. Further research will be conducted to find other approaches for comparison.

Acknowledgments. This research was funded by Faculty of Applied Science, King Mongkut's University of Technology North Bangkok. No. 651118.

References

1. Gerig, T.M., Sen, A.R.: MLE in two normal samples with equal but unknown population coefficients of variation. J. Am. Stat. Assoc. **75**, 704–708 (1980)
2. Searls, D.T.: The utilization of a known coefficient of variation in the estimation procedure. J. Am. Stat. Assoc. **59**, 1225–1226 (1964)
3. Niwitpong, S.: Confidence intervals for the normal mean with a known coefficient of variation. Far East J. Math. Sci. **97**, 711–727 (2015)
4. Srivastava, V.K.: On the use of coefficient of variation in estimating mean. J. Indian Soc. Agric. Stat. **26**, 33–36 (1974)
5. Srivastava, V.K.: A note on the estimation of mean in normal population. Metrika **27**(1), 99–102 (1980). https://doi.org/10.1007/BF01893580
6. Srivastava, V.K., Singh, R.S.: Uniformly minimum variance unbiased estimator of efficiency ratio in estimation of normal population mean. Statist. Probab. Lett. **10**, 241–245 (1990)
7. Sahai, A.: On an estimator of normal population mean and UMVU estimation of its relative efficiency. Appl. Math. Comput. **152**, 701–708 (2004)
8. Thangjai, W., Niwitpong, S., Niwitpong, S.: Confidence intervals for mean and difference of means of normal distributions with unknown coefficients of variation. Mathematics **5**, 1–23 (2017)
9. Thangjai, W., Niwitpong, S.-A., Niwitpong, S.: Bayesian confidence intervals for means of normal distributions with unknown coefficients of variation. In: Huynh, V.-N., Entani, T., Jeenanunta, C., Inuiguchi, M., Yenradee, P. (eds.) IUKM 2020. LNCS (LNAI), vol. 12482, pp. 361–371. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-62509-2_30
10. Krishnamoorthy, K., Lu, Y.: Inference on the common means of several normal populations based on the generalized variable method. Biometrics **59**, 237–247 (2003)
11. Lin, S.H., Lee, J.C.: Generalized inferences on the common mean of several normal populations. J. Stat. Plan. Inference **134**, 568–582 (2005)
12. Tian, L.: Inferences on the common coefficient of variation. Stat. Med. **24**, 2213–2220 (2005)
13. Tian, L., Wu, J.: Inferences on the common mean of several log-normal populations: the generalized variable approach. Biom. J. **49**, 944–951 (2007)
14. Thangjai, W., Niwitpong, S., Niwitpong, S.: Confidence intervals for the common mean of several normal populations. Stud. Computat. Intell. **692**, 321–331 (2017)

15. Jordan, S.J., Krishnamoorthy, K.: Exact confidence intervals for the common mean of several normal populations. Biometrics **52**, 77–86 (1996)
16. Fung, W.K., Tsang, T.S.: A simulation study comparing tests for the equality of coefficients of variation. Stat. Med. **17**, 2003–2014 (1998)
17. Thangjai, W., Niwitpong, S.: Confidence intervals for the weighted coefficients of variation of two-parameter exponential distributions. Cogent Math. **4**, 1–16 (2017)
18. Weerahandi, S.: Generalized confidence intervals. J. Am. Stat. Assoc. **88**, 899–905 (1993)
19. Ye, R.D., Ma, T.F., Wang, S.G.: Inferences on the common mean of several inverse Gaussian populations. Comput. Stat. Data Anal. **54**, 906–915 (2010)
20. Zou, G.Y., Donner, A.: Construction of confidence limits about effect measures: a general approach. Stat. Med. **27**, 1693–1702 (2008)
21. Zou, G.Y., Taleban, J., Hao, C.Y.: Confidence interval estimation for lognormal data with application to health economics. Comput. Stat. Data Anal. **53**, 3755–3764 (2009)
22. Graybill, F.A., Deal, R.B.: Combining unbiased estimators. Biometrics **15**, 543–550 (1959)