STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **7**, June 2019, pp 277–290.
Published online in International Academic Press (www.IAPress.org)

Performance of Some Confidence Intervals for Estimating the Population Coefficient of Variation Under both Symmetric and Skewed Distributions

Moustafa Omar Ahmed Abu-Shawiesh [1,*], Hayriye Esra Akyz [2], BM Golam Kibria [3]

[1] *Department of Mathematics, The Hashemite University, Jordan*
[2] *Department of Statistics, Bitlis Eren University, Turkey*
[3] *Department of Mathematics and Statistics, Florida International University, USA*

Abstract This paper aims to compare the performance of proposed confidence intervals for population coefficient of variation (CV) with the existing confidence intervals, namely, McKay, Miller, and Gulher et al. confidence intervals under both symmetric and skewed distributions. We observed that the proposed augmented-large-sample (AA&K-ALS) confidence interval performed well in terms of coverage probability in all cases. The large-sample (A&A-LS) and adjusted degrees of freedom (AA&K-ADJ) confidence intervals had much lower coverage probability than the nominal level for skewed distributions. However, the average widths of the AA&K-LS confidence interval are narrower than that of the rest confidence intervals. Two real-life data are analyzed to illustrate the implementation of the several methods.

Keywords Augmented Large Sample, Confidence Interval, Coefficient of Variation, Coverage Probability, Average Width.

AMS 2010 subject classifications 62F10, 62F35

DOI: 10.19139/soic.v7i2.630

1. Introduction

A *confidence interval* (CI) is a range of values that gives the user a sense of how precisely a statistic estimates a parameter. On the other hand, *coefficient of variation* (CV) also known as relative standard deviation, is a standardized measure of dispersion of a probability distribution or frequency distribution. It is a helpful quantity to describe the variation in evaluating results from different populations. It is also a dimensionless measure of the degree of variability relative to the mean. In statistical literature, the concept of CV was introduced by [22] and can be defined as a ratio of the population standard deviation to the population mean ($\mu \neq 0$) (or its absolute value, $|\mu|$) and given as follows:

$$CV = \frac{\sigma}{\mu} \tag{1}$$

The CV, as an important measure of variation, has been used in many fields such as medicine, biology, physics, finance, toxicology, business, engineering, life insurance and survival analysis, because it is free from the unit of measurement and it can be used for comparing the variability of two different populations. In practice, the population CV is unknown and needs to be estimated from data. To estimate the unknown population CV, one may consider either confidence interval or hypothesis testing. The confidence interval provides information respecting the population value of the quantity much more than the point estimate [5]. That is, confidence interval indicate that

*Correspondence to: Moustafa Omar Ahmed Abu-Shawiesh (Email: mabushawiesh@hu.edu.jo). Department of Mathematics, The Hashemite University. Al-Zarga, Jordan (13115).

ISSN 2310-5070 (online) ISSN 2311-004X (print)

the population parameter will be within this interval with a certain level of confidence as estimates for population parameters, while the hypothesis testing focuses on the use of statistical tests to accept or reject hypotheses concerning these parameters. The typical sample estimate of the population coefficient of variation (CV) is given as:

$$\hat{CV} = \frac{S}{\bar{X}} \tag{2}$$

where S is the sample standard deviation, the square root of the unbiased estimator of the population variance, and $\bar{X}$ is the sample mean. The point estimator of the population CV in (1) is a useful statistical measure, its confidence interval is more useful than the point estimator.

In this paper, we choose the CV as a parameter of our interest because of its widespread use in describing the variation within a data set. Moreover, among scale parameters, the CV is a more informative quantity than others. As noted in [8], the CV is preferred to the variance or standard deviation in various fields of interest, especially in biological and medical research.

The confidence interval for the CV given in literature is developed mostly based on the normality assumption. When the data are normally distributed, the *coverage probability* (CP) of this confidence interval is close to a nominal value of $1 - \alpha$. However, the underlying distributions are non-normal in many situations, like for example, the positively skewed data are common in real life, especially when sample sizes are small [1, 2, 3, 28]. In these situations, the CP of the confidence interval can be considerably below$1 - \alpha$. Hummel [12] presented a confidence interval for the population variance by adjusting the degrees of freedom of the chi-square distribution. In order to develop approximate confidence intervals for variance under non-normality, Burch [4] considered a number of kurtosis estimators combined with large-sample. There are various methods available for estimating the confidence interval for a population CV. For more information on the confidence interval for CV, we refer to [15, 18, 24, 17, 27, 16, 7, 20, 25, 10, 21] and recently [23] among others. The necessary sample size for estimating a population parameter is important. Therefore, determining the sample size to estimate the population CV is also important. Tables of necessary sample sizes to have sufficiently narrow confidence intervals under different scenarios are provided by Kelley [13].

The objective of this paper is to propose some new confidence intervals for estimating the population CV and compared them with some existing confidence intervals under the condition of symmetric and skewed distributions. A Monte-Carlo simulation will be conducted to compare the performance of the confidence intervals.

The rest of this paper is organized as follows: In Section 2, we review the confidence intervals for the variance under non-normality. In Section 3 , three important and useful existing confidence intervals for the population CV are reviewed. The proposed confidence intervals for the CV are presented in Section 4. To compare the performance of the interval estimators, a Monte-Carlo simulation study has been conducted in Section 5. Two real-life data are analyzed to illustrate the implementation of the several methods in Section 6. Finally, some concluding remarks are presented in Section 7.

2. The confidence intervals for the variance under non-normality

In this section, we review the confidence intervals for the variance under the non-normality assumption proposed by Hummel et. al. [12] and Burch [4].

2.1. The adjusted degrees of freedom confidence interval (ADJ)

Suppose, $X_1, X_2, X_3, \ldots, X_n \sim N(\mu, \sigma^2)$, then the sample variance, for samples sufficiently large, can be approximated as a chi-square with an appropriate estimate for the degrees of freedom. Hummel, et al. [12] using the method of matching moments found an estimate for the degrees of freedom, see for example [26]. They matched the first two moments of the distribution of the sample variance (S^2) and proposed the confidence interval for the population variance (σ^2) by adjusting the degrees (ADJ) of freedom of the chi-squared distribution. Hummel et al. [12] method, referred to as ADJ, has confidence interval limits (CI_{ADJ}) that are given as follows:

$$CI_{ADJ} = \left(\frac{\hat{r}\, S^2}{\chi^2_{\left(1-\frac{\alpha}{2}\,,\,\hat{r}\right)}} \;\;,\;\; \frac{\hat{r}\, S^2}{\chi^2_{\left(\frac{\alpha}{2}\,,\,\hat{r}\right)}} \right) \tag{3}$$

where $\chi^2_{\left(\frac{\alpha}{2}\,,\,\hat{r}\right)}$ and $\chi^2_{\left(1-\frac{\alpha}{2}\,,\,\hat{r}\right)}$ are the α/2 and 1$-\alpha$/2 quantiles of the central chi-squared distribution with $\hat{r}$ degrees of freedom, respectively where $\hat{r}$ is given as follows:

$$\hat{r} = \frac{2n}{\hat{\gamma} + \left[\frac{2n}{n-1}\right]} \tag{4}$$

$$\hat{\gamma} = \left[\frac{n\,(n+1)}{(n-1)\,(n-2)\,(n-3)} \, \frac{\sum_{i=1}^{n} \left(X_i - \bar{X}\right)^{\,4}}{S^{\,4}} \right] - \left[\frac{3\,(n-1)^2}{(n-2)\,(n-3)} \right] \tag{5}$$

where S^2 is the sample variance. If the random sample is known to come from a normal population, then $r = n - 1$ and Eq. (3) reduces to the classical chi-square confidence interval which will be given in the next section in Eq. (6).

2.2. The Large-Sample Confidence Interval for the Variance (LS)

Suppose, $X_1,\, X_2,\, X_3,\, \ldots,\, X_n \sim N(\mu,\, \sigma^2)$, then the $(1-\alpha)100\%$ confidence interval for the population variance using a pivotal quantity $Q = (n-1)\, S^2 / \sigma^2$, is referred to as CL [6], has confidence interval limits (CI$_{CL}$) given as follows:

$$CI_{CL} = \left(\frac{(n-1)\, S^2}{\chi^2_{\left(1-\frac{\alpha}{2}\,,\, n-1\right)}} \;\;,\;\; \frac{(n-1)\, S^2}{\chi^2_{\left(\frac{\alpha}{2}\,,\, n-1\right)}} \right) \tag{6}$$

where $\chi^2_{\left(\frac{\alpha}{2}\,,\, n-1\right)}$ and $\chi^2_{\left(1-\frac{\alpha}{2}\,,\, n-1\right)}$ are the α/2 and 1$-\alpha$/2 quantiles of the central chi-squared distribution with $n-1$ degrees of freedom, respectively. If the normality assumption is not valid, one can depend on large-sample (LS) theory which indicates that the sample variance is asymptotically normally distributed, that is:

$$S^2 \overset{Assymp}{\sim} N\left(\sigma^2,\, \frac{\sigma^4}{n}\left(\kappa_e + \frac{2n}{n-1}\right)\right) \tag{7}$$

where $\kappa_e = \frac{E\left[(X-\mu)^4\right]}{\left(E\left[(X-\mu)^2\right]\right)^2} - 3$ is the excess kurtosis of the distribution. In practice, a natural logarithm transformation of S^2 is applied in order to achieve approximate normality for the distribution of log(S^2) in a finite-sample applications. The mean and the variance of log(S^2) are estimated using the first two terms of a Taylor's series expansion implies that:

$$\log(S^2) \overset{Approx}{\sim} N\left(\log(\sigma^2),\, \frac{1}{n}\left(\kappa_e + \frac{2n}{n-1}\right)\right) \tag{8}$$

and therefore the $(1-\alpha)100\%$ large-sample confidence interval for the population variance (σ^2), referred to as LS, has confidence interval limits (*CI$_{LS}$*) given as follows:

$$CI_{LS} = \left(S^2 \,\exp\left(-Z_{1-\frac{\alpha}{2}}\sqrt{A}\right) \;\;,\;\; S^2 \,\exp\left(Z_{1-\frac{\alpha}{2}}\sqrt{A}\right) \right) \tag{9}$$

where $A = \frac{G_2 + 2n/(n-1)}{n}$, in this case κ_e has been replaced with the commonly used estimator G$_2$ defined by:

$$G_2 = \frac{n-1}{(n-2)(n-3)} \;\left[(n-1)\; g_2 + 6\right] \tag{10}$$

with $g_2 = \frac{m_4}{m_2^2} - 3$, $m_4 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^4}{n}$, *and* $m_2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n}$.

2.3. The Augmented-Large-Sample Confidence Interval for the Variance (ALS)

Burch [4] considered a modification to the approximate distribution of log(S^2) by using a three-term Taylor's series expansion. Employing the large-sample properties of S^2, the mean and the variance of log(S^2) are given by:

$$\begin{aligned} E\left(\log(S^2)\right) &\approx \log(\sigma^2) - \tfrac{1}{2n}\left(\kappa_e + \tfrac{2n}{n-1}\right) \\ Var\left(\log(S^2)\right) &\approx \tfrac{1}{n}\left(\kappa_e + \tfrac{2n}{n-1}\right)\left(1 + \tfrac{1}{2n}\left(\kappa_e + \tfrac{2n}{n-1}\right)\right) \end{aligned} \tag{11}$$

Both the mean and the variance of log(S^2) are dependent on the kurtosis of the underlying distribution and therefore the$(1-\alpha)100\%$ augmented-large-sample confidence interval for the population variance (σ^2), referred to as ALS, has confidence interval limits (CI_{ALS}) given as follows:

$$CI_{ALS} = \left(S^2 \exp\left(-Z_{1-\frac{\alpha}{2}}\sqrt{B} + C\right) \ ,\ S^2 \exp\left(Z_{1-\frac{\alpha}{2}}\sqrt{B} + C\right)\ \right) \tag{12}$$

where $B = V\hat{a}r\left(\log(S^2)\right)$, $C = \frac{\hat{\kappa}_{e\ ,\ 5} + 2n/(n-1)}{2n}$, in this case$\kappa_e$ has been replaced with the modified estimator $\hat{\kappa}_{e\ ,\ 5}$ defined by:

$$\hat{\kappa}_{e\ ,\ 5} = \left(\frac{n+1}{n-1}\right) G_2 \left(1 + \frac{5\,G_2}{n}\right) \tag{13}$$

3. The existing confidence interval for population coefficient of variation

In this section, three important and useful existing confidence intervals for the population CV are reviewed in order to compare them with the performance of the proposed methods in our paper.

3.1. McKay's Confidence Interval (McK)

McKay [17] developed a confidence interval for normal population CV by using the approximation method. McKay's method, referred to as McK, has confidence interval limits (CI_{McK}) givenas follows:

$$CI_{McK} = \left(\left(\frac{S}{\bar{X}}\right)\sqrt{\left(\frac{\chi^2_{v\ ,\ 1-\alpha/2}}{v+1} - 1\right)\left(\frac{S}{\bar{X}}\right)^2 + \frac{\chi^2_{v\ ,\ 1-\alpha/2}}{v}}\ ,\ \left(\frac{S}{\bar{X}}\right)\sqrt{\left(\frac{\chi^2_{v\ ,\ \alpha/2}}{v+1} - 1\right)\left(\frac{S}{\bar{X}}\right)^2 + \frac{\chi^2_{v\ ,\ \alpha/2}}{v}}\right) \tag{14}$$

where$\chi^2_{v\ ,\ 1-\alpha/2}$ and $\chi^2_{v\ ,\ \alpha/2}$ are respectively the 100((1$-\alpha$)/2) and 100(α/2) percentile of the chi-square distribution with $v = n-1$ degrees of freedom.

3.2. Miller's Confidence Interval (Mill)

Miller [18] proposed a confidence interval based on the sample coefficient of variation ($\hat{CV}$) that approximates an asymptotic normal distribution. Miller's method, referred to as Mill, has confidence interval limits (CI_{Mill}) given as follows:

$$CI_{Mill} = \left(\left(\frac{S}{\bar{X}}\right) - Z_{\,1-\alpha/2}\sqrt{\frac{(S/\bar{X})^2}{n-1}\left(0.5 + \left(\frac{S}{\bar{X}}\right)^2\right)}\ ,\ \left(\frac{S}{\bar{X}}\right) + Z_{\,1-\alpha/2}\sqrt{\frac{(S/\bar{X})^2}{n-1}\left(0.5 + \left(\frac{S}{\bar{X}}\right)^2\right)}\right) \tag{15}$$

where$Z_{\,1-\alpha/2}$ is the 100((1$-\alpha$)/2) percentile of the standard normal distribution.

3.3. Gulhar, Kibria, Albatineh & Ahmed's Confidence Interval (GKA&A)

Gulhar et al. [10] proposed a confidence interval for normal population CV based on the known formula for calculating the confidence interval for σ^2 given in Eq. (6), referred to as GKA&A, has confidence interval limits

($CI_{GKA\&A}$) given as follows:

$$CI_{GKA\&A} = \left(\frac{\sqrt{n-1}\,\left(S/\bar{X}\right)}{\sqrt{\chi^2_{v\ ,\ 1-\alpha/2}}}\ ,\ \frac{\sqrt{n-1}\,\left(S/\bar{X}\right)}{\sqrt{\chi^2_{v\ ,\ \alpha/2}}} \right) \tag{16}$$

where$\chi^2_{v\ ,\ 1-\alpha/2}$ and$\chi^2_{v\ ,\ \alpha/2}$are respectively the 100((1− α)/2) and 100(α/2) percentile of the chi-square distribution with $v = n-1$ degrees of freedom.

4. The proposed confidence intervals for population coefficient of variation

In this section, we propose three confidence intervals for estimating the population CV based on the confidence intervals for the population variance under non-normality. The first proposed confidence interval is based on Eq. (3), referred to as AA&K-ADJ. The second proposed confidence interval is based on Eq. (10), referred to as AA&K-LS and the third proposed confidence interval is based on Eq. (13), referred to as AA&K-ALS.

4.1. *Abu-Shawiesh, Akyüz & Kibria's Adjusted Degrees of Freedom Confidence Interval (AA&K-ADJ)*

From Eq. (3), we construct the confidence interval for the population CV based on the confidence interval for the population variance by adjusting the degrees of freedom of the chi-square distribution, which is:

$$P\left(\frac{\hat{r}\,S^2}{\chi^2_{\left(1-\frac{\alpha}{2}\ ,\ \hat{r}\right)}}\ <\ \sigma^2 < \ \frac{\hat{r}\,S^2}{\chi^2_{\left(\frac{\alpha}{2}\ ,\ \hat{r}\right)}} \right) = 1-\alpha \tag{17}$$

Assuming that$\mu \neq 0$, dividing this confidence interval by μ^2 results in

$$P\left(\frac{\hat{r}\,S^2}{\chi^2_{\left(1-\frac{\alpha}{2}\ ,\ \hat{r}\right)}\,\mu^2}\ <\ \left(\frac{\sigma}{\mu}\right)^2\ <\ \frac{\hat{r}\,S^2}{\chi^2_{\left(\frac{\alpha}{2}\ ,\ \hat{r}\right)}\,\mu^2} \right) = 1-\alpha \tag{18}$$

Since μ is not known, we can replace it by the unbiased estimator of μ which is resulting in

$$= P\left(\frac{\hat{r}}{\chi^2_{\left(1-\frac{\alpha}{2}\ ,\ \hat{r}\right)}}\widehat{CV}^2\ <\ CV^2\ <\ \frac{\hat{r}}{\chi^2_{\left(\frac{\alpha}{2}\ ,\ \hat{r}\right)}}\widehat{CV}^2 \right) = 1-\alpha \tag{19}$$

Taking the square root results in the final proposed confidence interval given by

$$= P\left(\sqrt{\frac{\hat{r}}{\chi^2_{\left(1-\frac{\alpha}{2}\ ,\ \hat{r}\right)}}}\widehat{CV}\ <\ CV\ <\ \sqrt{\frac{\hat{r}}{\chi^2_{\left(\frac{\alpha}{2}\ ,\ \hat{r}\right)}}}\widehat{CV} \right) = 1-\alpha \tag{20}$$

That is, the$(1-\alpha)100\%$ confidence interval for the population CV based on the confidence interval for the population variance (σ^2) by adjusting the degrees of freedom of the chi-square distribution is given by:

$$CI_{AA\&K-ADJ} = \left(\sqrt{\frac{\hat{r}}{\chi^2_{\left(1-\frac{\alpha}{2}\ ,\ \hat{r}\right)}}}\widehat{CV}\ \ ,\ \ \sqrt{\frac{\hat{r}}{\chi^2_{\left(\frac{\alpha}{2}\ ,\ \hat{r}\right)}}}\widehat{CV} \right) \tag{21}$$

4.2. *Abu-Shawiesh, Akyüz & Kibria's Large-Sample Confidence Interval (AA&K-LS)*

Similarly, from Eq. (9), we construct the confidence interval for the population coefficient of variation (CV) based on the large-sample confidence interval for the population variance which can be derived as follows:

$$P\left(S^2\ \exp\left(-Z_{1-\frac{\alpha}{2}}\sqrt{A}\right)\ <\ \sigma^2\ <\ S^2\ \exp\left(Z_{1-\frac{\alpha}{2}}\sqrt{A}\right) \right) = 1-\alpha \tag{22}$$

Assuming that $\mu \neq 0$, dividing this confidence interval by μ^2results in

$$P\left(\frac{S^2\,\exp\left(-Z_{1-\frac{\alpha}{2}}\sqrt{A}\right)}{\mu^2} < \left(\frac{\sigma}{\mu}\right)^2 < \frac{S^2\,\exp\left(Z_{1-\frac{\alpha}{2}}\sqrt{A}\right)}{\mu^2}\right) = 1-\alpha \tag{23}$$

Since μ is not known, we can replace it by the unbiased estimator of μ which is $\hat{\mu} = \bar{X}$ resulting in,

$$\begin{aligned} &P\left(\frac{S^2\,\exp\left(-Z_{1-\frac{\alpha}{2}}\sqrt{A}\right)}{\bar{X}^{\,2}} < CV^2 < \frac{S^2\,\exp\left(Z_{1-\frac{\alpha}{2}}\sqrt{A}\right)}{\bar{X}^{\,2}}\right) = 1-\alpha \\ &P\left(\overset{\wedge}{CV}{}^2\,\exp\left(-Z_{1-\frac{\alpha}{2}}\sqrt{A}\right) < CV^2 < \overset{\wedge}{CV}{}^2\,\exp\left(Z_{1-\frac{\alpha}{2}}\sqrt{A}\right)\right) = 1-\alpha \end{aligned} \tag{24}$$

Taking the square root results in the final proposed confidence interval given by

$$P\left(\overset{\wedge}{CV}\sqrt{\exp\left(-Z_{1-\frac{\alpha}{2}}\sqrt{A}\right)} < CV < \overset{\wedge}{CV}\sqrt{\exp\left(Z_{1-\frac{\alpha}{2}}\sqrt{A}\right)}\right) = 1-\alpha \tag{25}$$

That is, the$(1-\alpha)100\%$ confidence interval for the population coefficient of variation (CV) based on the large-sample confidence interval for the population variance (σ^2) is given by:

$$CI_{AA\&K-LS} = \left(\overset{\wedge}{CV}\sqrt{\exp\left(-Z_{1-\frac{\alpha}{2}}\sqrt{A}\right)},\ \overset{\wedge}{CV}\sqrt{\exp\left(Z_{1-\frac{\alpha}{2}}\sqrt{A}\right)}\right) \tag{26}$$

4.3. Abu-Shawiesh, Akyüz & Kibria's Augmented-Large-Sample Confidence Interval (AA&K-ALS)

Using Eq. (12), we construct the confidence interval for the population CV based on the augmented-large-sample confidence interval for the population variance which can be derived as follows:

$$P\left(S^2\,\exp\left(-Z_{1-\frac{\alpha}{2}}\sqrt{B}+C\right) < \sigma^2 < S^2\,\exp\left(Z_{1-\frac{\alpha}{2}}\sqrt{B}+C\right)\right) = 1-\alpha \tag{27}$$

Assuming that$\mu \neq 0$, dividing this confidence interval by μ^2results in

$$P\left(\frac{S^2\,\exp\left(-Z_{1-\frac{\alpha}{2}}\sqrt{B}+C\right)}{\mu^{\,2}} < \left(\frac{\sigma}{\mu}\right)^2 < \frac{S^2\,\exp\left(Z_{1-\frac{\alpha}{2}}\sqrt{B}+C\right)}{\mu^{\,2}}\right) = 1-\alpha \tag{28}$$

Since$\mu$is not known, we can replace it by the unbiased estimator of μwhich is $\hat{\mu} = \bar{X}$resulting in

$$P\left(\frac{S^2\,\exp\left(-Z_{1-\frac{\alpha}{2}}\sqrt{B}+C\right)}{\bar{X}^{\,2}} < CV^2 < \frac{S^2\,\exp\left(Z_{1-\frac{\alpha}{2}}\sqrt{B}+C\right)}{\bar{X}^{\,2}}\right) = 1-\alpha \tag{29}$$

$$P\left(\overset{\wedge}{CV}{}^2\exp\left(-Z_{1-\frac{\alpha}{2}}\sqrt{B}+C\right) < CV^2 < \overset{\wedge}{CV}{}^2\exp\left(Z_{1-\frac{\alpha}{2}}\sqrt{B}+C\right)\right) = 1-\alpha \tag{30}$$

Taking the square root results in the final proposed confidence interval given by

$$P\left(\overset{\wedge}{CV}\sqrt{\exp\left(-Z_{1-\frac{\alpha}{2}}\sqrt{B}+C\right)} < CV < \overset{\wedge}{CV}\sqrt{\exp\left(Z_{1-\frac{\alpha}{2}}\sqrt{B}+C\right)}\right) = 1-\alpha \tag{31}$$

That is, the$(1-\alpha)100\%$ confidence interval for the population CV based on the augmented-large-sample confidence interval for the population variance is given by:

$$CI_{AA\&K-ALS} = \left(\overset{\wedge}{CV}\sqrt{\exp\left(-Z_{1-\frac{\alpha}{2}}\sqrt{B}+C\right)},\ \overset{\wedge}{CV}\sqrt{\exp\left(Z_{1-\frac{\alpha}{2}}\sqrt{B}+C\right)}\right) \tag{32}$$

5. Simulation study

In this paper, we considered 6 confidence intervals, 3 existing intervals and 3 proposed intervals, for estimating the population CV and compared them under the same simulation conditions. A Monte-Carlo simulation is conducted using the statistical software MATLAB to compare the performance of the interval estimators. More on simuation study we refer our readers to Kibria and Banik [14], Gedam and Pathare [9] and very recently Muhammad, Tahani, and Frank [19] among others.The performance of the estimators considered for various CV values, sample sizes (n) and probability distributions.

5.1. Simulation technique

Random samples are generated from most commonly used distributions (both symmetric and skewed) with specific parameters, these distributions are:

1. Normal distribution with parameters $\mu = 10$ and $\sigma = 1,\ 3,\ 5,$
2. Chi-Square distribution with a parameter degrees of freedom (*df*) *v* = 200, 22, 8, and
3. Gamma distribution with parameters $\alpha = 100,\ 11.11,\ 4$ and $\beta = 2$.

The number of simulation replications was M = 50,000 for each case. The coefficient of variation and type I error were considered as CV = 0.10, 0.30, 0.50 and $\alpha = 0.05$, respectively. The CV was calculated for each one of the three distributions by utilizing the equations in Table 1.

Table 1. The CV and Skewness of data from Normal, Chi-Square and Gamma distributions

Distribution	CV	Skewness
Normal (μ, σ^2)	μ/σ	0
Chi-Square (*v*)	$\sqrt{2/v}$	$2\sqrt{2/v}$
Gamma(α, 2)	$1/\sqrt{\alpha}$	$2/\sqrt{\alpha}$

We will also obtain the $(1-\alpha)\,100\,\%$ confidence interval denoted by CI = (L, U) based on the 50,000 replicates and estimated the CP and the *average* width (AW), respectively, from the proportion of CIs containing the true CV over all MC simulations by using the following two formulas:

$$CoverageProbability(CP) = \frac{\#\,(L\ \leq\ CV\ \leq\ U)}{50000} \tag{33}$$

and

$$AverageWidth(AW) = \frac{\sum_{i=1}^{50000}(U_i\ -\ L_i)}{50000} \tag{34}$$

where $\#(L \leq CV \leq U)$ denotes the number of simulation runs for which the population CV lies within the confidence interval. The coverage probability is an excellent method for evaluating the success of a particular confidence interval in capturing the true parameter. The CP is calculated by counting the number of times the true CV is captured between the lower and upper limits. An interval width is calculated by subtracting a lower limit from an upper limit. A smaller width is better because it means that the true CV is captured within a smaller span and the results are more precise. The simulated coverage probabilities and average interval widths for normal, chi-square and gamma distributions are presented in Tables 2- 4, respectively.

Table 2. Estimated Coverage Probabilities and Average Widths of the 95% Confidence Intervals for the Normal Distribution

	n	Coefficient of Variation Confidence Interval					
		Existing Methods CP (AW)			Proposed Methods CP (AW)		
		McK	Mill	GKA&A	AA&K-ADJ	AA&K-LS	AA&K-ALS
CV = 0.10	15	0.9102	0.9131	0.9469	0.8379	0.9163	0.9499
		(0.0726)	(0.0735)	(0.0829)	(0.0556)	(0.0740)	(0.0805)
	25	0.9252	0.9266	0.9473	0.8933	0.9235	0.9493
		(0.0561)	(0.0565)	(0.0603)	(0.0489)	(0.0560)	(0.0593)
	50	0.9373	0.9382	0.9475	0.9219	0.9325	0.9489
		(0.0396)	(0.0397)	(0.0408)	(0.0366)	(0.0392)	(0.0404)
	100	0.9448	0.9453	0.9490	0.9356	0.9409	0.9475
		(0.0280)	(0.0280)	(0.0282)	(0.0267)	(0.0276)	(0.0279)
CV = 0.30	15	0.9112	0.9155	0.9298	0.8139	0.9018	0.9340
		(0.2404)	(0.2406)	(0.2508)	(0.1687)	(0.2244)	(0.2426)
	25	0.9263	0.9279	0.9294	0.8700	0.9061	0.9306
		(0.1839)	(0.1838)	(0.1818)	(0.1476)	(0.1690)	(0.1776)
	50	0.9381	0.9382	0.9286	0.8974	0.9131	0.9304
		(0.1290)	(0.1288)	(0.1228)	(0.1103)	(0.1181)	(0.1219)
	100	0.9447	0.9447	0.9306	0.9141	0.9217	0.9304
		(0.0909)	(0.0906)	(0.0849)	(0.0801)	(0.0831)	(0.0843)
CV = 0.50	15	0.9003	0.9070	0.8906	0.7627	0.8629	0.8961
		(0.4928)	(0.4641)	(0.4219)	(0.2873)	(0.3773)	(0.4088)
	25	0.9213	0.9261	0.8894	0.8181	0.8667	0.8923
		(0.3635)	(0.3514)	(0.3051)	(0.2474)	(0.2835)	(0.2994)
	50	0.9391	0.9392	0.8895	0.8513	0.8753	0.8927
		(0.2501)	(0.2439)	(0.2052)	(0.1844)	(0.1971)	(0.2039)
	100	0.9478	0.9446	0.8892	0.8693	0.8809	0.8902
		(0.1751)	(0.1712)	(0.1419)	(0.1341)	(0.1390)	(0.1413)

Table 3. Estimated Coverage Probabilities and Average Widths of the 95% Confidence Intervals for the Chi-Square Distribution

	n	Coefficient of Variation Confidence Interval					
		Existing Methods CP (AW)			Proposed Methods CP (AW)		
		McK	Mill	GKA&A	AA&K-ADJ	AA&K-LS	AA&K-ALS
CV = 0.10	15	0.9140	0.9161	0.9505	0.8459	0.9213	0.95384
		(0.0725)	(0.0734)	(0.0828)	(0.0558)	(0.0744)	(0.0810)
	25	0.9271	0.9291	0.9497	0.8972	0.9259	0.9523
		(0.0561)	(0.0565)	(0.0603)	(0.0488)	(0.0564)	(0.0598)
	50	0.9400	0.9408	0.9502	0.9250	0.93634	0.9547
		(0.0396)	(0.0397)	(0.0408)	(0.0368)	(0.0396)	(0.0411)
	100	0.9462	0.9465	0.9503	0.9375	0.9441	0.9535
		(0.0280)	(0.0280)	(0.0283)	(0.0268)	(0.0280)	(0.0285)
CV = 0.30	15	0.9177	0.9218	0.9463	0.8377	0.9234	0.9647
		(0.2382)	(0.2386)	(0.2491)	(0.1667)	(0.2321)	(0.2674)
	25	0.9351	0.9373	0.9447	0.8911	0.9305	0.9651
		(0.1837)	(0.1837)	(0.1817)	(0.1474)	(0.1791)	(0.1982)
	50	0.9472	0.9478	0.9411	0.9202	0.9403	0.9663
		(0.1293)	(0.1291)	(0.1230)	(0.1129)	(0.1280)	(0.1369)
	100	0.9532	0.9529	0.9406	0.9382	0.9487	0.9653
		(0.0913)	(0.0911)	(0.0852)	(0.0842)	(0.0919)	(0.0953)
CV = 0.50	15	0.9215	0.9283	0.9396	0.8233	0.9216	0.9748
		(0.4660)	(0.4454)	(0.4113)	(0.2720)	(0.4085)	(0.5287)
	25	0.9377	0.9420	0.9354	0.8800	0.9281	0.9773
		(0.3527)	(0.3424)	(0.3000)	(0.2450)	(0.3229)	(0.3843)
	50	0.9557	0.9566	0.9283	0.9179	0.9426	0.9773
		(0.2470)	(0.2411)	(0.2037)	(0.1943)	(0.2368)	(0.2634)
	100	0.9648	0.9637	0.9265	0.9437	0.9550	0.9789
		(0.1739)	(0.1701)	(0.1412)	(0.1505)	(0.1733)	(0.1839)

5.2. Results discussion

From Table 2, it is observed that coverage probabilities and average widths of proposed confidence intervals for Normal Distribution when CV = 0.10, 0.30, 0.50 and $\alpha = 0.05$ are very close to the nominal confidence level even for small sample sizes. The coverage probabilities of all confidence intervals are close to the nominal confidence level for each value of the coefficient of variation. Proposed three methods have performed very well in terms of average widths. As the value of the coefficient of variation decreases, the average width of confidence intervals decreases. It is seen that the proposed AA&K-ADJ confidence interval has narrowest average width compare to the rest of the interval estimators. On the other hand; as the value of the coefficient of variation decreased, narrower confidence intervals were obtained in all cases. As the coverage probabilities increase with increasing sample size, the average widths decrease. Similar results were observed when the distribution of the population was positively skewed. The coverage probabilities of proposed confidence intervals based on Chi-Square and Gamma distributions are quite close to the nominal confidence level for $\alpha = 0.05$. It is noted that the average widths of the confidence intervals are reduced as the sample size increases. The proposed confidence intervals performed as good as the other confidence intervals (Tables 3- 4).

Table 4. Estimated Coverage Probabilities and Average Widths of the 95% Confidence Intervals for the Gamma Distribution

	n	Coefficient of Variation Confidence Interval					
		Existing Methods CP (AW)			Proposed Methods CP (AW)		
		McK	Mill	GKA&A	AA&K-ADJ	AA&K-LS	AA&K-ALS
CV = 0.10	15	0.9105	0.9129	0.9499	0.8439	0.9185	0.9535
		(0.0727)	(0.0735)	(0.0830)	(0.0560)	(0.0745)	(0.0813)
	25	0.9290	0.9307	0.9501	0.8980	0.9276	0.9529
		(0.0561)	(0.0564)	(0.0603)	(0.0488)	(0.0564)	(0.0596)
	50	0.9397	0.9404	0.9485	0.9230	0.9357	0.9522
		(0.0396)	(0.0397)	(0.0408)	(0.0367)	(0.0396)	(0.0410)
	100	0.9455	0.9459	0.9498	0.9380	0.9430	0.95328
		(0.0280)	(0.0280)	(0.0282)	(0.0268)	(0.0280)	(0.0286)
CV = 0.30	15	0.9168	0.9215	0.9471	0.8384	0.9226	0.9647
		(0.2373)	(0.2377)	(0.2484)	(0.1663)	(0.2318)	(0.2666)
	25	0.9323	0.9344	0.9418	0.8888	0.9281	0.9644
		(0.1825)	(0.1825)	(0.1807)	(0.1471)	(0.1782)	(0.1990)
	50	0.9478	0.9485	0.9427	0.9211	0.9408	0.9654
		(0.1286)	(0.1284)	(0.1225)	(0.1121)	(0.1271)	(0.1356)
	100	0.9536	0.9534	0.9410	0.9381	0.9492	0.9647
		(0.0908)	(0.0905)	(0.0848)	(0.0837)	(0.0912)	(0.0944)
CV = 0.50	15	0.9215	0.9279	0.9419	0.8187	0.9213	0.9761
		(0.4645)	(0.4443)	(0.4106)	(0.2670)	(0.4074)	(0.5227)
	25	0.9384	0.9428	0.9368	0.8803	0.9271	0.9786
		(0.3526)	(0.3423)	(0.3000)	(0.2452)	(0.3228)	(0.3854)
	50	0.9546	0.9559	0.9305	0.9187	0.9418	0.9795
		(0.2470)	(0.2411)	(0.2037)	(0.1945)	(0.2367)	(0.2633)
	100	0.9644	0.9631	0.9256	0.9428	0.9551	0.9794
		(0.1738)	(0.1701)	(0.1412)	(0.1503)	(0.1728)	(0.1841)

6. Real Data

In this section, we consider two real-life examples to illustrate the performance of the proposed confidence intervals for the population coefficient of variation (CV).

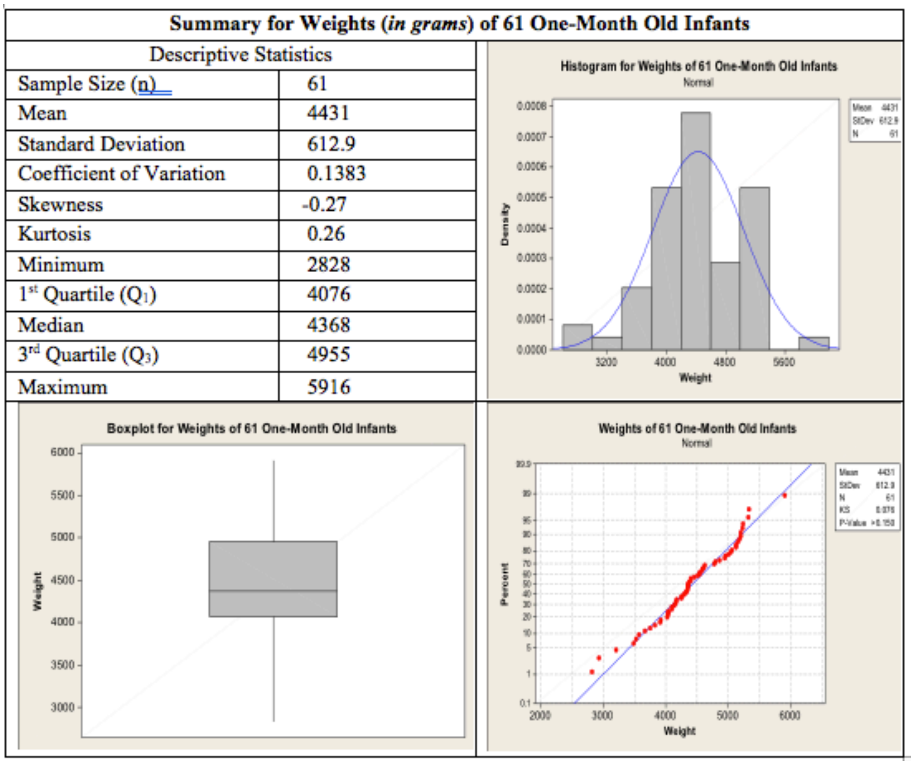
6.1. Example 1: Infants weights (in grams) data

The first data set was obtained from the study by Ziegler, Nelson, and Jeter [29]. The data represents the weights (*in grams*) of 61 one-month old infants listed as follows:

4960 5130 4260 5160 4050 5240 4350 4360 3930 4410 4610 4102 3530
4550 4460 2940 4160 4110 4410 4800 5130 3670 4550 4290 5210
4950 5210 3210 4030 3580 4360 4360 3920 4050 4630 3756 4382
4586 5336 2828 4172 4256 4594 4866 4784 4520 5238 4320 5070
5330 3836 5916 5010 4344 3496 4148 4044 5192 4368 4180 5044

A summary with descriptive statistics, Box-and-Whisker plot, the histogram, density plot, and normal probability plot from the data was obtained using Minitab® Release 14 (Minitab Inc.) and the results are shown in Figure 1.

As can be observed, the Kolmogorov-Smirnov (K-S) goodness-of-fit test for normality have a p-value greater than 0.05 (Shapiro-Wilk normality test, p-value=0.34), we conclude that the data are in excellent agreement with a normal distribution. Additionally, the histogram and the normal probability plot show a normal distribution. It appears from Kolmogorov-Smirnov (K-S) (p-value=0.38) that the given data follow a normal distribution with mean of 4500 and standard deviation of 615. Thus one may claim that the true CV of this data is, 0.14 (615/4500). The resulting 95% confidence interval and corresponding width for the proposed and existing confidence intervals of the population CV are calculated and reported in Table 5. From this table, we see that all the interval estimators captured the true CV, 0.14 and the average widths of the proposed confidence intervals performed as good as existing confidence intervals.



Descriptive Statistics	
Sample Size (n)	61
Mean	4431
Standard Deviation	612.9
Coefficient of Variation	0.1383
Skewness	-0.27
Kurtosis	0.26
Minimum	2828
1st Quartile (Q_1)	4076
Median	4368
3rd Quartile (Q_3)	4955
Maximum	5916

Figure 1. Descriptive statistics for the weights of 61 one-month old infants

Table 5. The 95% Confidence Intervals for the Population Coefficient of Variation of the Weight of One-Month Old Infants

Method	Confidence Interval Limits		
	Lower Limit	Upper Limit	Width
McK	0.1130	0.1633	0.0503
Mill	0.1130	0.1635	0.0505
GKA&A	0.1173	0.1683	0.0510
AA&K-ADJ	0.1163	0.1706	0.0543
AA&K-LS	0.1144	0.1672	0.0528
AA&K-ALS	0.1155	0.1687	0.0532

6.2. *Example 2: Postmortem Interval (PMI) Data*

The second data set was obtained from Banik and Kibria [25]. The data represents the postmortem interval (PMI) which is defined as the elapsed time between death and an autopsy. Knowledge of PMI is considered essential when conducting medical research on human cadavers. The following data are PMIs of 22 human brains pecimens obtained at autopsy in a recent study [11]:

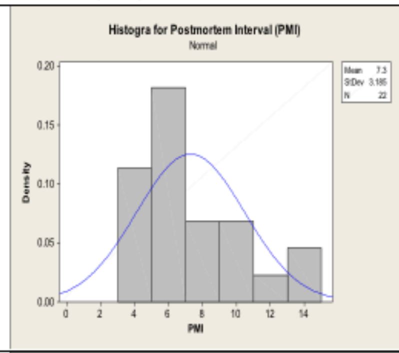
5.5, 14.5, 6, 5.5, 5.3, 5.8, 11.0, 6.1, 7.0, 14.5, 10.4,
4.6, 4.3, 7.2, 10.5, 6.5, 3.3, 7.0, 4.1, 6.2, 10.4, 4.9

A summary with descriptive statistics, Box-and-Whisker plot, the histogram, density plot, and normal probability plot from the data was obtained using Minitab® Release 14 (Minitab Inc.) and the results are shown in Figure 2.

According to Banik and Kibria [25] by using the Kolmogorov-Smirnov (K-S) goodness-of-fit test, the PMI data are from a gamma distribution with shape parameter, $\alpha = 5.25$, and scale parameter, $\beta = 1.39$. The population coefficient of variation $CV = \frac{\sigma}{\mu} = \frac{\sqrt{\alpha\beta}}{\alpha\beta} = \frac{1}{\sqrt{\alpha}} = \frac{1}{\sqrt{5.25}} = 0.4364$. The resulting 95% confidence interval and corresponding width for the proposed and existing intervals of the population coefficient of variation (CV) are calculated and reported in Table 6. From this table, we see that all the interval estimators contain the true CV, 0.4364 and the proposed AA&K-LS confidence interval performed better than other confidence intervals in the sense of smallest average width.

Summary for Postmortem Interval (PMI)

Descriptive Statistics	
Sample Size (n)	22
Mean	7.3
Standard Deviation	3.185
Coefficient of Variation	0.4363
Skewness	1.14
Kurtosis	0.48
Minimum	3.3
1st Quartile (Q_1)	5.2
Median	6.15
3rd Quartile (Q_3)	10.4
Maximum	14.5



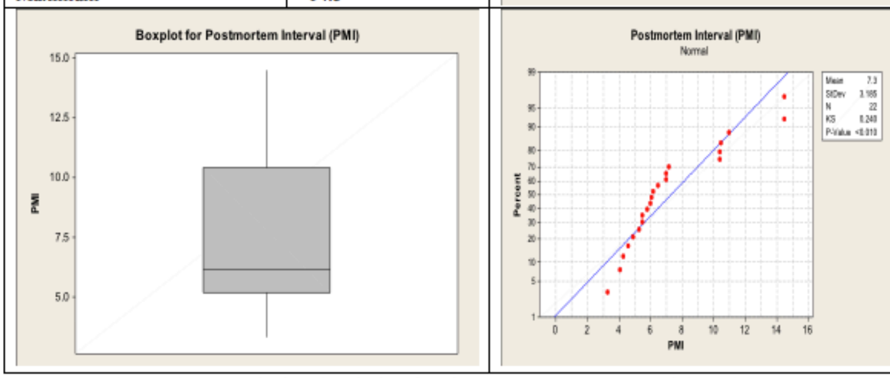


Figure 2. Summary for postmortem interval (PMI)

7. Summary and Concluding Remarks

In this paper, three confidence intervals based on variance were developed for the CV for both symmetric and skewed distributions. Since, a theoretical comparison among the estimators is not possible, a simulation study has been conducted to compare the performance of the estimators for all conditions. The large-sample (AA&K-LS) and adjusted degrees of freedom (AA&K-ADJ) confidence intervals had much lower coverage probability than the nominal level for skewed distributions. However, the average widths of AA&K-LS confidence interval are narrower

Table 6. The 95% Confidence Intervals for the Population Coefficient of Variation of the Postmortem Interval (PMI)

Method	Confidence Interval Limits		
	Lower Limit	Upper Limit	Width
McK	0.2718	0.5863	0.3145
Mill	0.2812	0.5913	0.3101
GKA&A	0.3356	0.6234	0.2878
AA&K-ADJ	0.3275	0.6532	0.3257
AA&K-LS	0.3122	0.6095	0.2973
AA&K-ALS	0.3128	0.6464	0.3336

than average widths of the others. In addition to the simulation, two real life data are analyzed for illustrating the findings of the paper which supported the findings of the simulation study of to some extent.

Acknowledgements

Authors are thankful to the referee for valuable comments and suggestions, which helped to improve the presentation of the paper.

REFERENCES

1. C. Almonte, and B. M. G. Kibria, *On some classical, bootstrap and transformation confidence intervals for estimating the mean of an asymmetrical population*, Model Assisted Statistics and Applications, vol. 4, no. 2, pp. 91–104, 2009.
2. P. G. Anderson, *Alternative confidence intervals for the total of a skewed biological population*, Ecology, vol. 85, no. 11, pp. 3166–3171, 2004.
3. B. Ayman, and B. M. G. Kibria, *One and two sample confidence intervals for estimating the mean of skewed populations: an empirical comparative study*, Journal of Applied Statistics, vol. 36, no. 6, pp. 601–609, 2009.
4. B. Burch, *Estimating kurtosis and confidence intervals for the variance under nonnormality*, Journal of Statistical Computation and Simulation, vol. 84, no. 12, pp. 2710–2720, 2014.
5. G. Casella, and R. L. Berger, *Statistical inference*, Thomson Learning, 2002.
6. V. Cosbasic, D. Loncar, *One-sided confidence intervals for population variances of skewed distributions*, Journal of Statistical Planning and Inference, vol. 141 no. 5, pp. 1667–1672, 2011.
7. J. D., Curto, and J. C. Pinto, *The coefficient of variation asymptotic distribution in the case of non-iid random variables*, Journal of Applied Statistics, vol. 36, no. 1, pp. 21–32, 2009.
8. A. Donner, and G. Y. Zou, *Closed-form confidence intervals for functions of the normal mean and standard deviation*, Statistical Methods in Medical Research, vol. 21, no. 4, pp. 347–359, 2012.
9. V. K. Gedam, and S. B. Pathare, *Approaches for mean response time of a two stage open queueing network model*, SOIC (Statistics, Optimization and Information Computing: An International Journal), vol. 3, no. 3, pp. 249–258, 2015.
10. M.Gulhar, B. M. G. Kibria, and A. N. Albatineh, *A comparison of some confidence intervals for estimating the population coefficient of variation: A simulation study*, Statistics and Operations Research Transactions, vol. 36, no. 1, pp. 45–68, 2012.
11. T. L. Hayes, and D. A. Lewis, *Better Confidence Intervals for the Variance in a Random Sample*, Brain and Language, vol.49, no. 3, pp. 289–308, 1995.
12. R. Hummel, S. Banga, and T. P. Hettmansperger, *Better confidence intervals for the variance in a random sample*, http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.482.4263.rep=rep1.type=pdf, 2005.
13. K.Kelley, *Sample size planning for the coefficient of variation from the accuracy in parameter estimation approach*, Behavior Research Methods, vol.39, no. 4, pp. 755–766, 2007.
14. B. M. G. Kibria, and S. Banik, *Parametric and nonparametric confidence intervals for estimating the difference of means of two skewed populations*, Journal of Applied Statistics, vol.40, no. 12, pp. 2617–2636, 2013.
15. L. H. Koopmans, D. B. Owen, and J. I. Rosenblatt, *Confidence intervals for the coefficient of variation for the normal and log normal distributions*, Biometrika, vol. 51, no. 1/2, pp. 25–32, 1964.
16. R. Mahmoudvand, and H. Hossein, *Two new confidence intervals for the coefficient of variation in a normal distribution*, Journal of Applied Statistics, vol. 36, no. 4, pp. 429–442, 2009.
17. A. T. McKay *Distribution of the coefficient of variation and the extended t distribution*, Journal of the Royal Statistical Society, vol. 95, no. 4, pp. 695–698, 1932.
18. G. E. Miller, *Asymptotic test statistics for coefficients of variation* Communications in Statistics - Theory and Methods, vol. 20, no. 10, pp. 3351–3363, 1991.
19. N. Muhammad, T. Coolen-Maturi, and F. P. Coolen, *Nonparametric predictive inference with parametric copulas for combining bivariate diagnostic tests* SOIC (Statistics, Optimization and Information Computing: An International Journal), vol. 6, no. 3, pp. 398–408, 2018.
20. W. Panichkitkosolkul, *Improved confidence intervals for a coefficient of variation of a normal distribution*, Thailand Statistician, vol. 7, no. 2, pp. 193–199, 2009.

21. W. Panichkitkosolkul, *Confidence intervals for the coefficient of variation in a normal distribution with a known population mean*, Journal of Probability and Statistics, vol. 7, no. 2, pp. 1–12, 2013.
22. K. Pearson, *Mathematical contributions to the theory of evolution III regression, heredity, and panmixia*, Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character, vol. 187, pp. 253–318, 1896.
23. P. Sangnawakij, and S. Niwitpong, *Confidence intervals for coefficients of variation in two-parameter exponential distributions*, Communications in Statistics - Simulation and Computation, vol. 46, pp. 6618–6630, 2017.
24. K. K. Sharma, and H. Krishna, *Asymptotic sampling distribution of inverse coefficient-of-variation and its applications*, IEEE Transactions on Reliability, vol. 43, no. 4, pp. 630–633, 1994.
25. B. Shipra, and B. M. G. Kibria, *Estimating the population coefficient of variation by confidence intervals*, Communications in Statistics - Simulation and Computation, vol. 40, no. 8, pp. 1236–1261, 2011.
26. L. H Shoemaker, *Fixing the F test for equal variances*, The American Statistician, vol. 57, no. 2, pp. 105–114, 2003.
27. M. G. Vangel, *Confidence intervals for a normal coefficient of variation*, The American Statistician, vol. 50, no. 1, pp. 21–26, 1996.
28. X. H. Zhou, and D. Phillip, *Nonparametric confidence intervals for the one- and two-sample problems*, Biostatistics, vol. 6, no. 2, pp. 187–200, 2005.
29. E. E. Ziegler, E. N. Steven, and M. J. Janice, *Iron Supplementation of Breastfed Infants*, Nutrition Reviews, vol. 69, pp. 71–77, 2011.