Adjusted Confidence Intervals for a Coefficient of Variation of a Normal Distribution

Wararit Panichkitkosolkul
Department of Mathematics and Statistics, Faculty of Science and Technology,
Thammasat University, Pathum Thani, 12121, Thailand.

Abstract

This paper presents a new confidence interval for a coefficient of variation of a normal distribution. The proposed confidence interval is constructed by adjusting the constant number in Vangel's [9] confidence interval. Monte Carlo simulations are used to investigate the behavior of this new confidence interval compared to the existing confidence intervals based on their coverage probabilities and expected lengths. Simulation results have shown that all cases of the new confidence interval have desired minimum coverage probabilities of 0.95 and 0.90. Moreover, this new one is better than the existing confidence intervals in terms of the expected length for all sample sizes and parameter values considered in this paper.

Keywords: confidence interval, coefficient of variation, coverage probability, expected length.

1. Introduction

The coefficient of variation is a dimensionless number that quantifies the degree of variability relative to the mean [1]. The population coefficient of variation is defined as:

$$\kappa = \frac{\sigma}{\mu}, \tag{1}$$

where σ is the population standard deviation and μ is the population mean. The typical sample estimate of κ is given as:

$$\hat{\kappa} = \frac{s}{\bar{x}}, \tag{2}$$

Where s is the sample standard deviation, the square root of the unbiased estimator of variance, and $\bar{x}$ is the sample mean.

The coefficient of variation has long been a widely used descriptive and inferential quantity in various areas of science, economics and others. In chemical experiments, the coefficient of variation is often used as a yardstick of precision of measurements; two measurement methods may be compared on the basis of their respective coefficients of variation. In finance, the coefficient of variation can be used as a measure of relative risks [2]. A test of the equality of the coefficients of variation for two stocks, can help to determine if the two stocks possess the same risk or not. Hamer et al. [3] used the coefficient of variation to assess homogeneity of bone test samples produced from a particular method to help assess the effect of external treatments, such as irradiation, on the properties of bones. Ahn [4] used the coefficient of variation in uncertainty analysis of fault trees. The coefficient of variation has also been employed by Gong and Li [5] in assessing the strength of ceramics.

Even though the estimated coefficient of variation can be a useful measure, perhaps the greatest use of it as a point estimate is to construct a confidence interval for the population quantity. A confidence interval provides much more information about the population value of the quantity of interest than does a point estimate (e.g., Smitson [6], Thompson [7], Steiger [8])

An approximate $(1-\alpha)100\%$ confidence interval for the coefficient of variation of a normal distribution (see, e.g., Vangel [9]) is given by:

$$CI = \left\{ \frac{\hat{\kappa}}{\sqrt{t_1(\theta_1\hat{\kappa}^2+1)-\hat{\kappa}^2}}, \frac{\hat{\kappa}}{\sqrt{t_2(\theta_2\hat{\kappa}^2+1)-\hat{\kappa}^2}} \right\} \tag{3}$$

where $\nu = n-1$, $t_1 = \chi^2_{\nu,1-\alpha/2}/\nu$, $t_2 = \chi^2_{\nu,\alpha/2}/\nu$ and $\theta = \theta(\nu,\alpha)$ is a known function selected so that a random variable $W_\nu = Y_\nu/\nu$, where Y_ν has a χ^2_ν distribution, has approximately the same distribution as a pivotal quantity $Q = \frac{K^2(1+\kappa^2)}{(1+\theta K^2)\kappa^2}$. This pivotal quantity can be used to construct hypothesis tests and a confidence interval for κ.

McKay [10] proposed that the choice $\theta = \frac{\nu}{\nu+1}$ gives a good approximation for the confidence interval in equation (3), but he was unable to investigate the small-sample distribution of Q. McKay's approximate confidence interval of a normal distribution is

$$CI_1 = \left\{ \begin{array}{l} \hat{\kappa}\left[\left(\frac{\chi^2_{\nu,1-\alpha/2}}{\nu+1}-1\right)\hat{\kappa}^2+\frac{\chi^2_{\nu,1-\alpha/2}}{\nu}\right]^{-1/2}, \\ \hat{\kappa}\left[\left(\frac{\chi^2_{\nu,\alpha/2}}{\nu+1}-1\right)\hat{\kappa}^2+\frac{\chi^2_{\nu,\alpha/2}}{\nu}\right]^{-1/2} \end{array} \right\} \tag{4}$$

where $\chi^2_{\nu,\alpha/2}$ and $\chi^2_{\nu,1-\alpha/2}$ are the $(\alpha/2)100\text{th}$ and $(1-\alpha/2)100\text{th}$ percentiles of the central chi-square distribution with $\nu = n-1$ degrees of freedom. Several authors have carried out numerical investigations of the accuracy of McKay's confidence interval. For instance, Iglewicz and Myers [11] compared McKay's confidence interval with the exact confidence interval based on the noncentral t distribution and they found that McKay's confidence interval is efficient for $n \geq 10$ and $0 < \kappa < 0.3$.

Vangel [9] proposed a new confidence interval for the coefficient of variation which he called the modified McKay's confidence interval. He proposed a choice for the function θ of $\theta = \frac{\nu}{\nu+1}\left[\frac{2}{\chi^2_{\nu,\alpha}}+1\right]$. He also suggested that the modified McKay method gave confidence intervals for the coefficient of variation that are closely related to the McKay's confidence interval but they are usually more accurate. The modified McKay's confidence interval, of a normal distribution is given by:

$$CI_2 = \left\{ \begin{array}{l} \hat{\kappa}\left[\left(\frac{\chi^2_{\nu,1-\alpha/2}+2}{\nu+1}-1\right)\hat{\kappa}^2+\frac{\chi^2_{\nu,1-\alpha/2}}{\nu}\right]^{-1/2}, \\ \hat{\kappa}\left[\left(\frac{\chi^2_{\nu,\alpha/2}+2}{\nu+1}-1\right)\hat{\kappa}^2+\frac{\chi^2_{\nu,\alpha/2}}{\nu}\right]^{-1/2} \end{array} \right\} \tag{5}$$

In this paper, we propose a new confidence interval for the coefficient of variation by adjusting the Vangel's confidence interval formula. The proposed confidence interval gives a shorter expected length. Additionally, we have compared coverage probabilities of this new confidence interval to the existing confidence intervals for a coefficient of variation.

The plan of the paper is as follows. Section 2 presents a proposed confidence interval for the coefficient of variation. Monte Carlo simulation results are given in Section 3. The conclusion is presented in Section 4.

2. A proposed confidence interval for the coefficient of variation

The adjustment of Vangel's [9] confidence interval by replacing the constant 1 in (5) with the positive constant number[1] c gives a proposed confidence interval for the coefficient of variation. This new confidence interval is given by:

$$CI_3 = \left\{ \begin{array}{l} \hat{\kappa}\left[\left(\dfrac{\chi^2_{\nu,1-\alpha/2}+2}{\nu+1}-c\right)\hat{\kappa}^2+\dfrac{\chi^2_{\nu,1-\alpha/2}}{\nu}\right]^{-1/2}, \\ \hat{\kappa}\left[\left(\dfrac{\chi^2_{\nu,\alpha/2}+2}{\nu+1}-c\right)\hat{\kappa}^2+\dfrac{\chi^2_{\nu,\alpha/2}}{\nu}\right]^{-1/2} \end{array} \right\} \quad (6)$$

It was found that if the positive constant number c is less than 1, both lower limit and upper limit are decreased. For this constant number adjustment, we may obtain a confidence interval for the coefficient of variation with shorter expected lengths and the proposed confidence interval also has minimum coverage probability $1-\alpha$.

[1] Using Monte Carlo simulations, if the constant c is a negative number, both lower limit and upper limit are increased. Moreover, the expected lengths of the proposed confidence interval are not shorter than that of the existing one. Therefore, we concentrate an only the positive constant number c.

The simulation results of the estimated coverage probabilities and expected lengths of the confidence intervals (4), (5) and (6) are presented in the next section.

3. Monte Carlo Simulations

In this section, we report the results of using Monte Carlo simulations to investigate the estimated coverage probabilities of the confidence intervals (4), (5) and (6) and their expected lengths. We used the R program [12, 13] to generate the data from normal distribution with κ = 0.1, 0.2 and 0.3, sample sizes; n = 10, 15, 25, 50 and 100. The number of simulation runs, M = 50,000 at level of significance α = 0.05 and 0.10. The positive constant number c in (6) has been selected so that coverage probabilities of the confidence intervals in (6) are at least $1-\alpha$ and close to $1-\alpha$. We found, using Monte Carlo simulation, that c = 0.9 is a good choice. Tables 1-2 show estimated coverage probabilities of the confidence intervals (4), (5) and (6), CI_1, CI_2 and CI_3, and their expected lengths at α = 0.05 and 0.10, respectively. As can be seen from Tables 1-2 and Figure 1, both confidence intervals (5) and (6), CI_2 and CI_3, have minimum coverage probability of $1-\alpha$ for all sample sizes and values of κ. In addition, the new confidence interval, CI_3, gives slightly higher coverage probabilities than the confidence interval CI_2. However the coverage probabilities of CI_1 in (4) are less than $1-\alpha$ in some situations. Furthermore, the expected lengths of CI_3 shown in Figure 2 are shorter than that of CI_1 and CI_2 in all conditions. For other choices of the constant c in (6), i.e. c = 0.95 and 1.1, the estimated coverage probabilities and expected lengths of CI_1, CI_2 and CI_3 are shown in Tables 3-4, respectively. These results show that c = 0.95 and 1.1 are not suitable choices for CI_3 because some estimated coverage probabilities are less than 0.95.

Table1. The estimated coverage probabilities and expected lengths of 95% confidence intervals in (4), (5) and (6) for a normal distribution where c = 0.9 in (6).

n	κ	Coverage probabilities			Expected lengths		
		CI_1	CI_2	CI_3	CI_1	CI_2	CI_3
10	0.1	0.9505	0.9505	0.9505	0.1132	0.1126	0.1123
	0.2	0.9493	0.9500	0.9503	0.2455	0.2392	0.2363
	0.3	0.9506	0.9516	0.9524	0.4328	0.4011	0.3877
15	0.1	0.9504	0.9506	0.9506	0.0844	0.0841	0.0839
	0.2	0.9496	0.9502	0.9505	0.1789	0.1766	0.1750
	0.3	0.9495	0.9507	0.9513	0.2966	0.2872	0.2806
25	0.1	0.9518	0.9519	0.9520	0.0612	0.0611	0.0610
	0.2	0.9507	0.9511	0.9515	0.1280	0.1272	0.1262
	0.3	0.9509	0.9513	0.9516	0.2065	0.2034	0.1998
50	0.1	0.9504	0.9505	0.9505	0.0414	0.0413	0.0413
	0.2	0.9501	0.9503	0.9508	0.0858	0.0855	0.0849
	0.3	0.9503	0.9509	0.9515	0.1364	0.1356	0.1335
100	0.1	0.9501	0.9502	0.9501	0.0286	0.0286	0.0285
	0.2	0.9509	0.9511	0.9515	0.0591	0.0590	0.0586
	0.3	0.9508	0.9510	0.9510	0.0934	0.0931	0.0918

Table2. The estimated coverage probabilities and expected lengths of 90% confidence intervals in (4), (5) and (6) for a normal distribution where c = 0.9 in (6).

n	κ	Coverage probabilities			Expected lengths		
		CI_1	CI_2	CI_3	CI_1	CI_2	CI_3
10	0.1	0.9011	0.9015	0.9016	0.0909	0.0904	0.0902
	0.2	0.8989	0.9004	0.9008	0.1945	0.1903	0.1883
	0.3	0.8984	0.9010	0.9018	0.3315	0.3123	0.3039
15	0.1	0.9003	0.9003	0.9013	0.0691	0.0689	0.0682
	0.2	0.9019	0.9027	0.9037	0.1451	0.1434	0.1375
	0.3	0.9001	0.9013	0.9022	0.2377	0.2310	0.2263
25	0.1	0.8999	0.9000	0.9001	0.0506	0.0505	0.0500
	0.2	0.9023	0.9024	0.9030	0.1054	0.1048	0.1040
	0.3	0.9002	0.9008	0.9013	0.1694	0.1670	0.1642
50	0.1	0.9002	0.9002	0.9001	0.0345	0.0344	0.0344
	0.2	0.8999	0.9002	0.9018	0.0713	0.0711	0.0689
	0.3	0.9007	0.9011	0.9016	0.1131	0.1125	0.1108
100	0.1	0.9005	0.9005	0.9006	0.0239	0.0239	0.0239
	0.2	0.9008	0.9013	0.9013	0.0494	0.0493	0.0490
	0.3	0.9010	0.9013	0.9011	0.0779	0.0777	0.0766

Table3. The estimated coverage probabilities and expected lengths of 95% confidence intervals in (4), (5) and (6) for a normal distribution where $c = 0.95$ in (6).

n	κ	Coverage probabilities			Expected lengths		
		CI_1	CI_2	CI_3	CI_1	CI_2	CI_3
10	0.1	0.9487	0.9489	0.9489	0.1135	0.1128	0.1127
	0.2	0.9496	0.9498	0.9500	0.2455	0.2393	0.2378
	0.3	0.9484	0.9497	0.9499	0.4368	0.4037	0.3967
15	0.1	0.9479	0.9481	0.9481	0.0846	0.0843	0.0842
	0.2	0.9501	0.9505	0.9507	0.1785	0.1762	0.1754
	0.3	0.9505	0.9507	0.9508	0.2960	0.2867	0.2833
25	0.1	0.9499	0.9500	0.9500	0.0613	0.0612	0.0611
	0.2	0.9497	0.9498	0.9499	0.1278	0.1270	0.1265
	0.3	0.9502	0.9511	0.9513	0.2065	0.2035	0.2016
50	0.1	0.9505	0.9506	0.9507	0.0414	0.0413	0.0413
	0.2	0.9494	0.9494	0.9496	0.0857	0.0855	0.0852
	0.3	0.9499	0.9501	0.9502	0.1363	0.1354	0.1344
100	0.1	0.9515	0.9515	0.9516	0.0286	0.0286	0.0286
	0.2	0.9498	0.9498	0.9499	0.0591	0.0590	0.0589
	0.3	0.9519	0.9522	0.9517	0.0933	0.0931	0.0924

Table4. The estimated coverage probabilities and expected lengths of 95% confidence intervals in (4), (5) and (6) for a normal distribution where $c = 1.1$ in (6).

n	κ	Coverage probabilities			Expected lengths		
		CI_1	CI_2	CI_3	CI_1	CI_2	CI_3
10	0.1	0.9500	0.9504	0.9503	0.1136	0.1129	0.1133
	0.2	0.9499	0.9502	0.9499	0.2451	0.2389	0.2419
	0.3	0.9505	0.9511	0.9508	0.4327	0.4005	0.4155
15	0.1	0.9519	0.9520	0.9519	0.0845	0.0843	0.0844
	0.2	0.9498	0.9502	0.9500	0.1787	0.1765	0.1782
	0.3	0.9510	0.9512	0.9510	0.2962	0.2868	0.2938
25	0.1	0.9510	0.9511	0.9510	0.0613	0.0612	0.0613
	0.2	0.9505	0.9509	0.9504	0.1280	0.1272	0.1282
	0.3	0.9503	0.9507	0.9500	0.2068	0.2037	0.2075
50	0.1	0.9502	0.9503	0.9502	0.0413	0.0413	0.0414
	0.2	0.9496	0.9500	0.9494	0.0858	0.0855	0.0861
	0.3	0.9510	0.9512	0.9507	0.1363	0.1355	0.1376
100	0.1	0.9502	0.9502	0.9500	0.0286	0.0286	0.0286
	0.2	0.9514	0.9514	0.9511	0.0591	0.0591	0.0594
	0.3	0.9518	0.9520	0.9508	0.0934	0.0931	0.0945

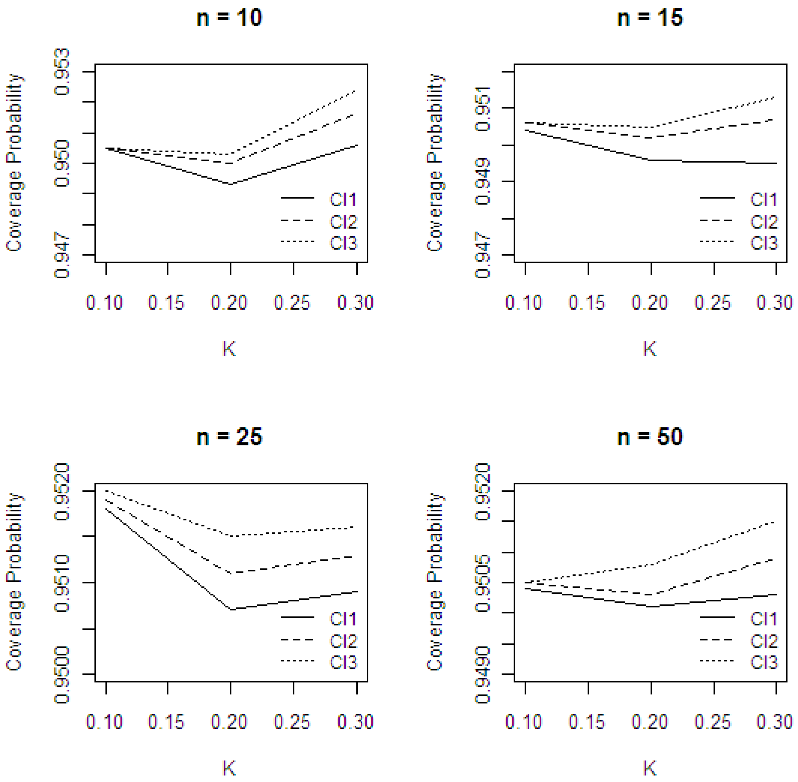
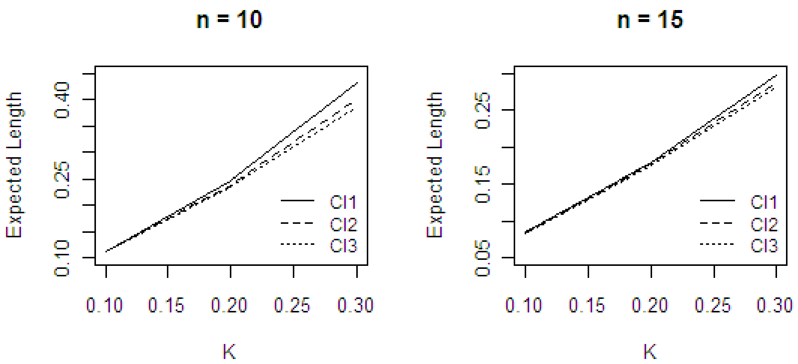


Figure 1. The estimated coverage probabilities of 95% confidence intervals, CI_1, CI_2 and CI_3 where $c = 0.9$ in (6)



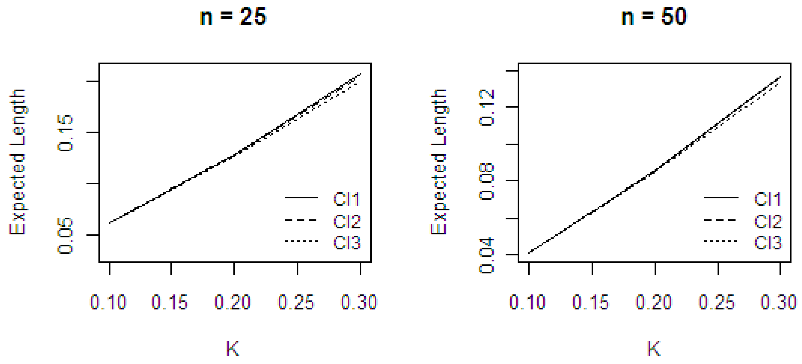


Figure 2. The estimated expected lengths of 95% confidence intervals, CI_1, CI_2 and CI_3 where $c = 0.9$ in (6)

4. Conclusion

We have proposed a new confidence interval for the coefficient of variation of a normal distribution. The McKay's [10] confidence interval, Vangel's [9] confidence interval and the proposed confidence interval are compared in this study. The new confidence interval is based on a constant number adjustment in the Vangel's confidence interval. The Vangel's confidence interval and the new confidence interval have minimum coverage probabilities $1-\alpha$. This new confidence interval performs better than the McKay's confidence interval and Vangel's confidence interval in terms of the expected length. Therefore, the proposed confidence interval is preferable to the existing confidence intervals since it has a shorter expected length along with the designed coverage probabilities for all sample sizes and values of κ considered here.

5. Acknowledgments

The author would like to thank two anonymous referees and Dr. Gareth Clayton for numerous helpful comments and suggestions, leading to many improvements in the paper.

6. References

[1] Kelley, K., 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.

[2] Miller, E.G., Karson, M.J., Testing the Equality of Two Coefficients of Variation, American Statistical Association: Proceedings of the Business and Economics Section, Part I, pp.278-283, 1977.

[3] Hamer, A.J., Strachan, J.R., Black, M.M., Ibbotson, C., Elson, R.A., A New Method of Comparative Bone Strength Measurement, Journal of Medical Engineering and Technology, Vol.19, No.1, pp.1-5, 1995.

[4] Ahn, K., Use of Coefficient of Variation for Uncertainty Analysis in Fault Tree Analysis, Reliability Engineering and System Safety, Vol.47, No.3, pp.229-230, 1995.

[5] Gong, J., Li, Y., Relationship between the Estimated Weibull Modulus and the Coefficient of Variation of the Measured Strength for Ceramics, Journal of the American Ceramic Society, Vol.82, No.2, pp.449-452, 1999.

[6] Smithson, M., Correct Confidence Intervals for Various Regression Effect Sizes and Parameters: The Importance of Noncentral Distributions in Computing Intervals, Educational and Psychological Measurement, Vol.61, pp.605-632, 2001.

[7] Thompson, B., What Future Quantitative Social Science Research Could Look Like: Confidence Intervals for Effect Sizes, Educational Researcher, Vol.31, pp.25-32, 2002.

[8] Steiger, J. H., Beyond the F test: Effect Size Confidence Intervals and Tests of Close Fit in the Analysis of Variance and Contrast Analysis, Psychological Methods, Vol.9, pp.164-182, 2004.

[9] Vangel, M.G. Confidence Intervals for a Normal Coefficient of Variation, The American Statistician, Vol.15, pp.21-26, 1996.

[10] McKay, A.T., Distribution of the Coefficient of Variation and the Extended t Distribution, Journal of the Royal Statistics Society, Vol.95, pp.695-698, 1932.

[11] Iglewicz, B., Myers, R.H., Comparisons of Approximations to the Percentage Points of the Sample Coefficient of Variation, Technometrics, Vol.12, pp.166-169, 1970.

[12] R Development Core Team, An Introduction to R, R Foundation for Statistical Computing, Vienna, 2008a.

[13] R Development Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, 2008b.