Thailand Statistician
July 2009; 7(2) : 193-199
www.statassoc.or.th
Contributed paper

Improved 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, Phathum Thani, 12121, Thailand.
E-mail: wararit@mathstat.sci.tu.ac.th

Received: 1 May 2009
Accepted: 11 July 2009.

Abstract

This paper presents a new confidence interval for a coefficient of variation of a normal distribution. The proposed confidence interval is constructed by replacing the typical sample coefficient of variation in Vangel's confidence interval with the maximum likelihood estimator. Monte Carlo simulation is 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: coefficient of variation, confidence interval, 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] and 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 (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 $\nu = n-1$, $t_1 \equiv \chi^2_{\nu,1-\alpha/2}/\nu$, $t_2 \equiv \chi^2_{\nu,\alpha/2}/\nu$ and $\theta = \theta(\nu,\alpha)$ is a known function selected so that a random variable $W_\nu \equiv Y_\nu/\nu$, where Y_ν has a χ^2_ν distribution, has approximately the same distribution as a pivotal quantity $Q \equiv \frac{K^2(1+\kappa^2)}{(1+\theta K^2)\kappa^2}$. This pivotal quantity can be used to construct hypothesis tests and confidence interval for κ.

McKay [10] proposed that the choice $\theta = \dfrac{\nu}{\nu+1}$ 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 is

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

where $\nu = n-1$ is degrees of freedom of χ^2 distribution. Several authors have carried out numerical investigations of the accuracy of McKay's confidence interval. For instance, Iglewicz and Myers [11] had 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 the choice for the function θ by $\theta = \dfrac{\nu}{\nu+1}\left[\dfrac{2}{\chi^2_{\nu,\alpha}}+1\right]$. 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 for a coefficient of variation is given by

$$CI_2 = \left\{ \hat{\kappa}\left[\left(\frac{\chi^2_{\nu,1-\alpha/2}+2}{\nu+1}-1\right)\hat{\kappa}^2+\frac{\chi^2_{\nu,1-\alpha/2}}{\nu}\right]^{-1/2}, \hat{\kappa}\left[\left(\frac{\chi^2_{\nu,\alpha/2}+2}{\nu+1}-1\right)\hat{\kappa}^2+\frac{\chi^2_{\nu,\alpha/2}}{\nu}\right]^{-1/2} \right\} \qquad (5).$$

Singh [12] said that the classical sample estimate of population coefficient of variation, $\hat{\kappa}$ is biased for κ. Furthermore, Mahmoudvand et al. [13] showed that the maximum likelihood estimator of population coefficient of variation is better than the sample coefficient of variation since the sample coefficient of variation may lie outside the bounds obtained for the population coefficient of variation. Also the maximum likelihood estimators are consistent estimators of their parameters and asymptotically efficient [14]. Therefore, in this paper, we propose the new confidence interval for the coefficient of variation of normal distribution by replacing the sample coefficient of variation in Vangel's confidence interval with the maximum likelihood estimator [9]. 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 confidence interval [9] by replacing the sample coefficient of variation, $\hat{\kappa}$, in (5) with the maximum likelihood estimator gives a proposed confidence interval for the coefficient of variation. The maximum likelihood estimator of population coefficient of variation for normal distribution is defined by

$$\tilde{\kappa} = \frac{\left(\sum_{i=1}^{n}(x_i-\bar{x})^2\right)^{1/2}}{\sqrt{n}\,\bar{x}}.$$

Thus, the new confidence interval for the coefficient of variation is given by

$$CI_3 = \left\{\tilde{\kappa}\left[\left(\frac{\chi^2_{\nu,1-\alpha/2}+2}{\nu+1}-1\right)\tilde{\kappa}^2+\frac{\chi^2_{\nu,1-\alpha/2}}{\nu}\right]^{-1/2},\ \tilde{\kappa}\left[\left(\frac{\chi^2_{\nu,\alpha/2}+2}{\nu+1}-1\right)\tilde{\kappa}^2+\frac{\chi^2_{\nu,\alpha/2}}{\nu}\right]^{-1/2}\right\} \quad (6).$$

In the next section, we present the simulation results, using Monte Carlo simulation, to estimate coverage probabilities and expected lengths of the confidence intervals (4), (5) and (6).

3. Monte Carlo simulation

In this section, we report the results of using Monte Carlo simulation to investigate the estimated coverage probabilities of the confidence intervals (4), (5) and (6) and their expected lengths. We used R program [15,16] 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. 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 for a normal distribution at α = 0.05 and 0.10, respectively. As can be seen from Tables 1-2, 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, almost all 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 are shorter than that of CI_1 and CI_2 in all conditions.

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

n	κ	Coverage probabilities			Expected lengths		
		CI_1	CI_2	CI_3	CI_1	CI_2	CI_3
10	0.1	0.9513	0.9514	0.9536	0.1133	0.1127	0.1067
	0.2	0.9497	0.9502	0.9513	0.2453	0.2390	0.2250
	0.3	0.9495	0.9508	0.9506	0.4342	0.4015	0.3732
15	0.1	0.9499	0.9500	0.9505	0.0845	0.0842	0.0813
	0.2	0.9498	0.9502	0.9512	0.1789	0.1766	0.1700
	0.3	0.9498	0.9506	0.9510	0.2965	0.2871	0.2748
25	0.1	0.9520	0.9520	0.9526	0.0612	0.0612	0.0599
	0.2	0.9495	0.9500	0.9506	0.1280	0.1272	0.1244
	0.3	0.9499	0.9507	0.9509	0.2064	0.2034	0.1984
50	0.1	0.9499	0.9500	0.9507	0.0414	0.0413	0.0409
	0.2	0.9513	0.9512	0.9518	0.0857	0.0855	0.0846
	0.3	0.9500	0.9503	0.9501	0.1364	0.1355	0.1339
100	0.1	0.9509	0.9508	0.9511	0.0286	0.0286	0.0284
	0.2	0.9499	0.9500	0.9506	0.0591	0.0591	0.0587
	0.3	0.9516	0.9517	0.9518	0.0935	0.0932	0.0927

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

n	κ	Coverage probabilities			Expected lengths		
		CI_1	CI_2	CI_3	CI_1	CI_2	CI_3
10	0.1	0.9001	0.9004	0.9039	0.0909	0.0904	0.0856
	0.2	0.9011	0.9016	0.9027	0.1942	0.1901	0.1792
	0.3	0.8991	0.9013	0.9024	0.3311	0.3121	0.2916
15	0.1	0.8996	0.8995	0.9018	0.0690	0.0688	0.0664
	0.2	0.8981	0.8985	0.9010	0.1450	0.1433	0.1380
	0.3	0.9010	0.9018	0.9016	0.2371	0.2305	0.2209
25	0.1	0.8984	0.8985	0.9011	0.0506	0.0505	0.0495
	0.2	0.8982	0.8989	0.9006	0.1054	0.1048	0.1025
	0.3	0.8994	0.8999	0.9002	0.1692	0.1668	0.1628
50	0.1	0.9007	0.9009	0.9019	0.0345	0.0344	0.0341
	0.2	0.9010	0.9010	0.9016	0.0713	0.0711	0.0703
	0.3	0.9025	0.9029	0.9029	0.1132	0.1125	0.1112
100	0.1	0.9001	0.9001	0.9012	0.0239	0.0239	0.0238
	0.2	0.9006	0.9006	0.9003	0.0494	0.0493	0.0490
	0.3	0.9022	0.9021	0.9027	0.0779	0.0777	0.0772

4. Conclusion

We have proposed a new confidence interval for the coefficient of variation. The McKay's confidence interval [10], Vangel's confidence interval [9] and the proposed confidence interval are compared in this study. The new confidence interval is based on the replacement the classical sample coefficient of variation in Vangel's confidence interval with the maximum likelihood estimator. The Vangel's confidence interval and the new confidence interval has 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 for all sample sizes and values of κ considered here.

Acknowledgements

The author would like to thank anonymous referees and Dr.Gareth Clayton for his helpful comments.

References

[1] Kelley, K. Sample size planning for the coefficient of variation from the accuracy in parameter estimation approach. *Behavior Research Methods*, 2007;39:755-766.

[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*, 1977;278-283.

[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*, 1995;19:1-5.

[4] Ahn, K. Use of coefficient of variation for uncertainty analysis in fault tree analysis, *Reliability Engineering and System Safety*, 1995;47:229-230.

[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*, 1999;82:449-452.

[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*, 2001;61:605-632.

[7] Thompson, B. What future quantitative social science research could look like: Confidence intervals for effect sizes. *Educational Researcher*, 2002;31:25-32.

[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*, 2004;9:164-182.

[9] Vangel, M.G. Confidence intervals for a normal coefficient of variation. *The American Statistician*, 1996;15:21-26.

[10] McKay, A.T. Distribution of the coefficient of variation and the extended t distribution. *Journal of the Royal Statistics Society*, 1932;95:695-698.

[11] Iglewicz, B., Myers, R.H. Comparisons of approximations to the percentage points of the sample coefficient of variation. *Technometrics*, 1970;12:166-169.

[12] Singh, M. Behaviour of sample coefficient of variation drawn from several distributions. *Sankhya*, 1993; 55: 65-76.

[13] Mahmoudvand, R., Hassani, H., Wilson, R. Is the sample coefficient of variation a good estimator for population coefficient of variation?. *World Applied Science Journal*, 2007; 2: 519-522.

[14] Casella, G., Berger, R.L. Statistical Inference. CA: Duxbury, 2002.

[15] R Development Core Team. *An introduction to R*. Vienna: R Foundation for Statistical Computing, 2008a.

[16] R Development Core Team. *R: A language and environment for statistical computing*, Vienna: R Foundation for Statistical Computing, 2008b.