World Academy of Science, Engineering and Technology
International Journal of Mathematical and Computational Sciences
Vol:7, No:9, 2013

On Simple Confidence Intervals for the Normal Mean with a Known Coefficient of Variation

Suparat Niwitpong, Sa-aat Niwitpong

Abstract—In this paper we proposed the new confidence interval for the normal population mean with a known coefficient of variation. In practice, this situation occurs normally in environment and agriculture sciences where we know the standard deviation is proportional to the mean. As a result, the coefficient of variation of is known. We propose the new confidence interval based on the best unbiased estimator for the population mean in the sense of minimum variance and this new confidence interval will compare with the existing confidence interval. We derive analytic expressions for the coverage probability and the expected length of each confidence interval. A numerical method will be used to assess the performance of these intervals based on their expected lengths.

Keywords—Confidence interval, coverage probability, expected length, known coefficient of variation.

I. Introduction

The confidence interval for the normal population mean has been studied recently. In the routine text such as Walpole et al. [8] shows that the confidence intervals for the normal means can be constructed in two cases; a) variance is known b) variance is unknown. However, in practice, there are situations in area of environmental and physical sciences that coefficients of variation are known. For example, in environmental studies, Bhat and Rao [1] argued that there are some situations that show the standard deviation of a pollutant is directly related to the mean, that means the τ is known. Furthermore in clinical chemistry, " when the batches of some substance (chemicals) are to be analyzed, if sufficient batches of the substances are analyzed, their coefficients of variation will be known". Brazauskas and Ghorai [2] also gave some examples in medical, biological and chemical experiments shown that in practice there are problems concerning that coefficients of variation are known. Most of this statistical problem is due to the estimation of the mean of normal distribution with known coefficient of variation see e.g. Khan [3] , Searls ([6], [7]) and the references cited in the mentioned papers. Recent paper of Bhat and Rao [1] extended the mentioned papers to the test for the normal mean, when its coefficient of variation is known. Their simulation results shown that, for the two-sided alternatives the likelihood ratio test or the Wald test is the best test. Niwitpong [5] extended the recent work of Bhat and Rao [1] to the confidence interval for the normal population mean with a known coefficient of variation. She proposed two new confidence intervals based on the estimator of the mean with known coefficient of variation of Searls [6] and the confidence interval constructed based on the relation between the coefficient of variation and the population mean and it standard deviation that will show in the next section. In this paper, we also proposed the new confidence interval for the mean of normal population mean with a known coefficient of variation based on the best unbiased estimator for the mean with a known coefficient of variation based on Khan [3]. We assess these three confidence intervals using the coverage probabilities and their expected lengths. Typically, we prefer confidence interval with coverage probability at least the nominal value $(1-\alpha)$ and its expected length is short.

Suparat Niwitpong and Sa-aat Niwitpong are with the Department of Applied Statistics, King Mongkut's University of Technology North Bangkok, Bangkok, THAILAND, 10800, (e-mail: suparatn@kmutnb.ac.th).

II. Confidence intervals for normal population mean

Let $X = [X_1, \ldots, X_n]$ be a random sample from the normal distribution with mean μ and standard deviation σ. The sample mean and variance for X are, respectively, denoted as $\bar{X}$ and S^2 when $\bar{X} = n^{-1}\sum_{i=1}^{n} X_i$, and $S^2 = (n-1)^{-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$. We are interested in 100(1-α)% confidence interval for μ. In practice, σ^2 is an unknown parameter and S^2 is an unbiased estimator for σ^2. In this case, a well-known $100(1-\alpha)\%$ confidence for μ is

$$CI_\mu = \left[\bar{X} - c\frac{S}{\sqrt{n}}, \bar{X} + c\frac{S}{\sqrt{n}}\right] \tag{1}$$

where c is $t_{1-\alpha/2}$, an upper $1-\alpha/2$ percentiles of the t-distribution with n-1 degrees of freedom. In the next section we review a method to construct a confidence interval for μ with known coefficient of variation.

III. Confidence intervals for normal population mean with a known coefficient of variation

It is known that the unbiased estimator for μ is the mean $\bar{X}$, however, when we known a prior information, for example the coefficient of variation ($\tau = \sigma/\mu$), the estimator $\bar{X}$ is not appropriate in terms of the mean square error. Searls [6] proposed the estimator $\bar{X}^* = (n+\tau^2)^{-1}\sum_{i=1}^{n} X_i$ where he showed that this estimator has lower mean squares error than that of the unbiased estimator. Further, it is easy to show that the variance of the estimator $\bar{X}^*$ is $\frac{nS^2}{(n+\tau^2)^2}$. Hence, a $100(1-\alpha)\%$ confidence for μ with known coefficient of variation τ is

$$CI_s = \left[\bar{X}^* - d\sqrt{\frac{nS^2}{(n+\tau^2)^2}}, \bar{X}^* + d\sqrt{\frac{nS^2}{(n+\tau^2)^2}}\right] \tag{2}$$

where d is $z_{1-\alpha/2}$, an upper $1-\alpha/2$ percentiles of the standard normal-distribution.

We now propose the new confidence interval for μ based on a prior information $\tau = \sigma/\mu$. It is easy to see that $\mu = \sigma/\tau$, hence our proposed confidence interval for μ is equal to the confidence interval of σ/τ which is

$$CI_p = \left[\frac{1}{\tau}\sqrt{\frac{(n-1)S^2}{u}}, \frac{1}{\tau}\sqrt{\frac{(n-1)S^2}{v}} \right] \quad (3)$$

where $u \sim \chi^2_{(n-1),(1-\alpha/2)}$ and $v \sim \chi^2_{(n-1),\alpha/2}$ and u and v are an upper $1-\alpha/2$ percentiles and lower $\alpha/2$ percentiles of the Chi-squared distribution.

In 1968, Khan [3] proposed the best unbiased estimator for μ in the sense of minimum variance with its variance when τ is known. We therefore construct the confidence interval for μ, using the best unbiased estimator (see, e.g. Khan [3], p. 1009), which is

$$CI_b = \left[\hat{\delta} - d\sqrt{\frac{\tau^2 S_n^2}{n(1+2\tau^2)}}, \hat{\delta} + d\sqrt{\frac{\tau^2 S_n^2}{n(1+2\tau^2)}} \right] \quad (4)$$

where where $\hat{\delta} = \alpha\delta_2 + (1-\alpha)\delta_1$, $0 < \alpha < 1$ and $\alpha = \tau^2/(\tau^2 + n(k_1/k_2 - 1))$, $k_1 = (n-1)\Gamma^2((n-1)/2)$, $k_2 = 2\Gamma^2(n/2)$, $\delta_1 = \bar{X}$, $\delta_2 = c_n\sqrt{n}S_n$ and $S_n^2 = n^{-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$.

Note that confidence intervals (2), (3) and (4) have a priori information τ in their confidence intervals whereas confidence interval (1) has no a priori information in its confidence interval. Niwitpong [5] reported that confidence interval CI_μ is not appropriate to used when τ is known. She also reported that confidence intervals CI_s and CI_p are shorter than that of confidence interval CI_μ when τ is known. In this paper, we therefore, consider confidence intervals CI_s and CI_p compared to our proposed confidence interval CI_b. Like Niwitpong [5], we compare all proposed confidence intervals based on their coverage probabilities and their expected lengths.

IV. Coverage probabilities and Expected length of each confidence interval

Following Niwitpong and Niwitpong [4], we now derive an analytic expression for the coverage probability for CI_μ. Let $P(\mu \in CI_\mu)$ be the coverage probability of confidence interval CI_μ then an analytic expression for this confidence coverage is given in Theorem 1, below,

Theorem 1 [Niwitpong [5]] The coverage probability and expected length for CI_μ are respectively

$E[\Phi(A_2) - \Phi(A_1)]$, $A_1 = \frac{-cS}{\sigma}$, $A_2 = \frac{cS}{\sigma}$, Z is the standard normal distribution and $\Phi(\cdot)$ is a standard normal function and

$$E(CI_\mu) = \frac{2c}{\sqrt{n}}E(s) = \frac{2c}{\sqrt{n}}\frac{\sqrt{2}\Gamma(n/2)}{\sqrt{n-1}\Gamma((n-1)/2)}\sigma$$

Proof See Niwitpong [5].

Similarly to CI_μ, we now derive an analytic expression for the coverage probability for CI_s. Let $P(\mu \in CI_s)$ be the coverage probability of confidence interval CI_s then an analytic expression for this confidence coverage is given in Theorem 2, below,

Theorem 2 [Niwitpong [5]] The coverage probability and expected length for CI_s are respectively

$E[\Phi(B_2) - \Phi(B_1)]$, $B_1 = \frac{-d}{W}\sqrt{\frac{nS^2}{(n+\tau^2)^2}}$, $B_2 = \frac{d}{W}\sqrt{\frac{nS^2}{(n+\tau^2)^2}}$, $W = \sqrt{\frac{n\sigma^2}{(n+\tau^2)^2}}$, and

$$E(CI_s) = 2d\sqrt{\frac{n}{(n+\tau^2)^2}}\frac{\sqrt{2}\Gamma(n/2)}{\sqrt{n-1}\Gamma((n-1)/2)}\sigma$$

Proof See Niwitpong [5].

Similarly to CI_μ, we now derive an analytic expression for the coverage probability for CI_p. Let $P(\mu \in CI_p)$ be the coverage probability of confidence interval CI_p then an analytic expression for this confidence coverage is given in Theorem 3.

Theorem 3 [Niwitpong [5]] The coverage probability and expected length for CI_p are respectively

$E[\Phi(C_2) - \Phi(C_1)]$, where $C_1 = \frac{\bar{X}}{Q} - \frac{1}{\tau Q}\sqrt{\frac{(n-1)S^2}{v}}$, $C_2 = \frac{\bar{X}}{Q} - \frac{1}{\tau Q}\sqrt{\frac{(n-1)S^2}{u}}$, $Q = \sigma/\sqrt{n}$ and

$$\begin{aligned} E(CI_p) &= \frac{1}{\tau}\sqrt{\frac{(n-1)}{v}}\frac{\sqrt{2}\Gamma(n/2)}{\sqrt{n-1}\Gamma((n-1)/2)}\sigma \\ &- \frac{1}{\tau}\sqrt{\frac{(n-1)}{u}}\frac{\sqrt{2}\Gamma(n/2)}{\sqrt{n-1}\Gamma((n-1)/2)}\sigma. \end{aligned}$$

Proof See Niwitpong [5].

Similarly to CI_μ, we now derive an analytic expression for the coverage probability for CI_b. Let $P(\mu \in CI_b)$ be the coverage probability of confidence interval CI_b then an analytic expression for this confidence coverage is given in Theorem 4.

Theorem 4 The coverage probability and expected length for CI_b are respectively

$E[\Phi(C_4) - \Phi(-C_4)]$, where $C_4 = \frac{d}{C_3}\sqrt{\frac{\tau^2 S_n^2}{n(1+2\tau^2)}}$, $C_3 = \sqrt{\frac{\tau^2\sigma^2}{n(1+2\tau^2)}}$ and

$$E(CI_b) = \left(\frac{2d}{C_3}\sqrt{\frac{\tau^2}{n(1+2\tau^2)}}\right)\frac{\sqrt{n}\Gamma((n-1)/2)}{\sqrt{2\tau^2}\Gamma(n/2)}\sigma.$$

Proof Similarly to Niwitpong [5],

$$\begin{aligned} & P(\mu \in CI_b) \\ = \;& P\left[\hat{\delta} - d\sqrt{\frac{\tau^2 S_n^2}{n(1+2\tau^2)}} < \mu < \hat{\delta} + d\sqrt{\frac{\tau^2 S_n^2}{n(1+2\tau^2)}}\right] \\ = \;& P\left[-d\sqrt{\frac{\tau^2 S_n^2}{n(1+2\tau^2)}} < \mu - \hat{\delta} < +d\sqrt{\frac{\tau^2 S_n^2}{n(1+2\tau^2)}}\right] \\ = \;& P\left[\frac{-d}{C_3}\sqrt{\frac{\tau^2 S_n^2}{n(1+2\tau^2)}} < \frac{\hat{\delta} - \mu}{C_3} < \frac{d}{C_3}\sqrt{\frac{\tau^2 S_n^2}{n(1+2\tau^2)}}\right] \\ = \;& P[-C_4 < Z < C_4] \\ = \;& E[I_{\{-C_4 < Z < C_4\}}(\tau)] \end{aligned}$$

$$
\begin{aligned}
&= E[E[I_{\{-C_4<Z<C_4\}}(\tau)]|S^2] \\
&= E[\Phi(C_4)-\Phi(-C_4)]
\end{aligned}
$$

where

$$
I_{\{-C_4<Z<C_4\}}(\tau)=\begin{cases} 1, & \text{if } \tau\in\{-C_4<Z<C_4\} \\ 0, & \text{otherwise}\end{cases}
$$

$C_3=\sqrt{\frac{\tau^2\sigma^2}{n(1+2\tau^2)}}$, $C_4=\frac{d}{C_3}\sqrt{\frac{\tau^2S_n^2}{n(1+2\tau^2)}}$.

The expected length of CI_b is therefore

$$
\begin{aligned}
&E(CI_b) \\
&= E\left(\frac{2d}{C_3}\sqrt{\frac{\tau^2S_n^2}{n(1+2\tau^2)}}\right) \\
&= \left(\frac{2d}{C_3}\sqrt{\frac{\tau^2}{n(1+2\tau^2)}}\right)E(S_n) \\
&= \left(\frac{2d}{C_3}\sqrt{\frac{\tau^2}{n(1+2\tau^2)}}\right)\frac{\sqrt{n}\Gamma((n-1)/2)}{\sqrt{2\tau^2}\Gamma(n/2)}\sigma.
\end{aligned}
$$

V. Simulation framework

In this section, since we have proved that the coverage probabilities of all confidences intervals are equal to $1-\alpha$, we therefore only use numerical method to assess three confidence intervals notified in the previous section: CI_b, CI_s and CI_p based on their average length widths. We design our experiment, without losing generality, by setting $\sigma=1$, $\tau=0, 0.01, 0.02, 0.03, 0.04, 0.05, 0.5, 0.1, 0.3, 0.5, 0.8, 0.9, 1, 1.1, 1.3, 1.5, 1.9, 2.5$ and the samples sizes $n=10, 30, 50, 100$. We wrote function in R program to generate the data which is normally distributed with zero mean and unit variance and then compute the average length width of each confidence interval: CI_b, CI_s and CI_p. All results are illustrated in Table I.

VI. Simulation results

From Table I, we found that the expected length of the confidence interval CI_b with a known coefficient of variation, $\tau\leq 0.10$ when $n=10$,$\tau\leq 0.03$ when $n=30$, $\tau\leq 0.02$ when $n=50$ and $\tau\leq 0.01$ when $n=100$, is shorter than other confidence intervals. The confidence interval CI_s is shorter than the confidence interval CI_p when $\tau\leq 0.1, n=10$,$\tau\leq 0.5, n=30, 50, 100$ otherwise the confidence interval CI_p is preferable.

TABLE I
The expected lengths of CI_b, CI_s and CI_p

τ	$E(CI_b)$	$E(CI_s)$	$E(CI_p)$	$E(CI_b)$	$E(CI_s)$	$E(CI_p)$
		$n=10$			$n=30$	
0.00	-	1.2744	-	-	0.7218	-
0.01	0.1143	1.2744	116.975	0.2092	0.7218	55.2651
0.02	0.2287	1.2744	58.487	0.4185	0.7218	27.6325
0.03	0.3431	1.2744	38.991	0.6278	0.7218	18.4217
0.04	0.4575	1.2744	29.243	0.8371	0.7218	13.8162
0.05	0.5719	1.2741	23.3951	1.0464	0.7218	11.0530
0.10	1.1438	1.2731	11.6975	2.0928	0.7216	5.5265
0.30	3.4314	1.2630	3.8991	6.2784	0.7197	1.8421
0.50	5.7191	1.2433	2.3395	10.464	0.7159	1.1053
0.80	9.1506	1.1977	1.4621	16.742	0.7067	0.6908
0.90	10.294	1.1789	1.2997	18.835	0.7028	0.6140
1.00	11.438	1.1585	1.1697	20.928	0.6985	0.5526
1.10	12.582	1.1368	1.0634	23.021	0.6938	0.5024
1.30	14.869	1.0901	0.8998	27.206	0.6833	0.4251
1.50	17.157	1.0403	0.7798	31.392	0.6715	0.3684
1.90	21.732	0.9363	0.6156	39.763	0.6443	0.2908
2.50	28.595	0.7842	0.4679	52.320	0.5974	0.2210
		$n=50$			$n=100$	
0.00	-	0.5571	-	-	0.3978	-
0.01	0.2729	0.5571	41.2900	0.3890	0.3929	28.7385
0.02	0.5459	0.5571	20.6450	0.7780	0.3929	14.2192
0.03	0.8189	0.5571	13.7633	1.1671	0.3929	9.4795
0.04	1.0919	0.5571	10.3225	1.5561	0.3929	7.1096
0.05	1.3649	0.5571	8.2580	1.9452	0.3929	5.6877
0.10	2.7299	0.5570	4.1290	3.8904	0.3929	2.8438
0.30	8.1899	0.5561	1.3763	11.6713	0.3926	0.9479
0.50	13.649	0.5544	0.8258	19.4522	0.3920	0.5687
0.80	21.839	0.5501	0.5161	31.1235	0.3904	0.3554
0.90	24.569	0.5483	0.4587	35.0139	0.3898	0.3159
1.00	27.299	0.5462	0.4190	38.9044	0.3890	0.2843
1.10	30.029	0.5440	0.3753	42.7948	0.3882	0.2585
1.30	35.489	0.5389	0.3189	50.5757	0.3864	0.2187
1.50	40.949	0.5332	0.2752	58.3566	0.3843	0.1895
1.90	51.869	0.5196	0.2173	73.9184	0.3792	0.1496
2.50	68.249	0.4952	0.1651	97.2610	0.3698	0.1137

VII. Conclusion

In this paper we proposed new confidence intervals, CI_b, for the normal population mean with a known coefficient of variation based on the best unbiased estimator of Khan [3]. This new confidence interval will compare to our previous confidence intervals, see e.g. Niwitpong [5]. We derived, mathematically, coverage probabilities and expected lengths of these intervals. It is shown in sections IV that the coverage probabilities of CI_b and CI_s and CI_p are equal to the nominal value $1-\alpha$. From numerical results in Table I, it is recommended that for a known small coefficient of variation and a small sample size, our new confidence interval CI_b is preferable. The confidence interval CI_s outperforms other confidence intervals when $\tau\leq 0.5$ otherwise we choose the confidence interval CI_p.

Acknowledgment

The author is grateful to King Mongkut's University of Technology North Bangkok for its finance support.

References

[1] K. Bhat and K. A. Rao, On Tests for a Normal Mean with Known Coefficient of Variation, International Statistical Review , 75 (2007), 170-182

[2] V. Brazauskas and J. Ghorai, Estimating the common parameter of normal models with known coefficients of variation: a sensitivity study of asymptotically efficient estimators, Journal of Statistical Computation and Simulation, 77(2007), 663-681

[3] R. A. Khan, A Note on Estimating the Mean of a Normal Distribution with Known Coefficient of Variation, Journal of the American Statistical Association, 63(1968), 1039-1041.

[4] S. Niwitpong and S. Niwitpong, Confidence interval for the difference of two normal population means with a known ratio of variances, Applied Mathematical Sciences , 4 (2010), 347 - 359.

[5] S. Niwitpong, Confidence interval for the normal means with known a coefficient of variation, World Academic of Science, Engineering and Technology 69 (2012), 677 - 680.
[6] D. T. Searls, A Note on the Use of an Approximately Known Coefficient of Variation, The American Statistician, 21(1967), 20-21.
[7] D. T. Searls, The Utilization of a Known Coefficient of Variation in the Estimation Procedure, Journal of the American Statistical Association, 59(1964), 1225-1226.
[8] R.E. Walpole, R. H. Myers, S.L. Myers, K. Ye., Probability & Statistics for Engineers & Scientists, Prentice Hall, New Jersey, 2002.