World Academy of Science, Engineering and Technology
International Journal of Mathematical and Computational Sciences
Vol:6, No:9, 2012

Confidence intervals for the normal mean with known coefficient of variation

Suparat Niwitpong

Abstract—In this paper we proposed two new confidence intervals for the normal population mean with known coefficient of variation. This situation occurs normally in environment and agriculture experiments where the scientist knows the coefficient of variation of their experiments. We propose two new confidence intervals for this problem based on the recent work of Searls [5] and the new method proposed in this paper for the first time. We derive analytic expressions for the coverage probability and the expected length of each confidence interval. Monte Carlo simulation 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 problem of inference concerning the confidence interval for the normal population mean has been studied recently. In the routine text such as Walpole et al. [7] 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 agricultural, biological, environmental and physical sciences that a 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 ([5], [6]) 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. This paper extends the recent work of Bhat and Rao [1] to the confidence interval for the normal population mean with known coefficient of variation. We propose two new confidence intervals based on the estimator of the mean with known coefficient of variation of Searls [5] where he showed that the mean squares error of his proposed estimator is lower than the unbiased estimator (the sample mean). We construct this confidence interval using the central limit theorem. The other confidence interval is constructed base on the relation between the coefficient of variation and the population mean and it standard deviation that will show in the next section. We assess these two new confidence intervals using the coverage probability and the ratio of expected lengths between confidence intervals. Typically, we prefer confidence interval with coverage probability at least the nominal value $(1-\alpha)$ and its expected length is short.

Suparat Niwitpong is 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.

1) Standard 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 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 appropriated. Searls [5] 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

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

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.

Note that confidence intervals (2) and (3) have a priori information τ in their confidence intervals whereas confidence interval (1) has no a priori information in its confidence interval.

IV. Coverage probabilities and Expected length of each confidence interval for mean

2) Coverage probability and Expected length of CI_μ: 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 by

$$\begin{aligned} & P(\mu \in CI_\mu) \\ = & P\left[\bar{X} - c\frac{S}{\sqrt{n}} < \mu < \bar{X} + c\frac{S}{\sqrt{n}}\right] \\ = & P\left[-c\frac{S}{\sqrt{n}} < \mu - \bar{X} < c\frac{S}{\sqrt{n}}\right] \\ = & P\left[\frac{-c\frac{S}{\sqrt{n}}}{\frac{\sigma}{\sqrt{n}}} < \frac{\mu - \bar{X}}{\frac{\sigma}{\sqrt{n}}} < \frac{c\frac{S}{\sqrt{n}}}{\frac{\sigma}{\sqrt{n}}}\right] \\ = & P\left[\frac{-cS}{\sigma} < Z < \frac{cS}{\sigma}\right] \\ = & P[A_1 < Z < A_2] \\ = & E[I_{\{A_1<Z<A_2\}}(\tau)] \\ = & E[E[I_{\{A_1<Z<A_2\}}(\tau)]|S^2] \\ = & E[\Phi(A_2) - \Phi(A_1)] \end{aligned}$$

where

$$I_{\{A_1<Z<A_2\}}(\tau) = \begin{cases} 1, & \text{if } \tau \in \{A_1 < Z < A_2\} \\ 0, & \text{otherwise} \end{cases}$$

$A_1 = \frac{-cS}{\sigma}$, $A_2 = \frac{cS}{\sigma}$, Z is the standard normal distribution and $\Phi(\cdot)$ is a standard normal function. The expected length of CI_μ is therefore

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

3) Coverage probability and Expected length of CI_s: 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 by

$$\begin{aligned} & P(\mu \in CI_s) \\ = & P\left[\bar{X}^* - d\sqrt{\frac{nS^2}{(n+\tau^2)^2}} < \mu < \bar{X}^* + d\sqrt{\frac{nS^2}{(n+\tau^2)^2}}\right] \\ = & P\left[-d\sqrt{\frac{nS^2}{(n+\tau^2)^2}} < \mu - \bar{X}^* < d\sqrt{\frac{nS^2}{(n+\tau^2)^2}}\right] \\ = & P\left[\frac{-d}{W}\sqrt{\frac{nS^2}{(n+\tau^2)^2}} < \mu - \bar{X}^* < \frac{d}{W}\sqrt{\frac{nS^2}{(n+\tau^2)^2}}\right] \\ = & P[B_1 < Z < B_2] \\ = & E[I_{\{B_1<Z<B_2\}}(\tau)] \\ = & E[E[I_{\{B_1<Z<B_2\}}(\tau)]|S^2] \\ = & E[\Phi(B_2) - \Phi(B_1)] \end{aligned}$$

where

$$I_{\{B_1<Z<B_2\}}(\tau) = \begin{cases} 1, & \text{if } \tau \in \{B_1 < Z < B_2\} \\ 0, & \text{otherwise} \end{cases}$$

$B_1 = \frac{-d}{W}\sqrt{\frac{nS^2}{(n+\tau^2)^2}}$, $B_2 = \frac{d}{W}\sqrt{\frac{nS^2}{(n+\tau^2)^2}}$ and $W = \sqrt{\frac{n\sigma^2}{(n+\tau^2)^2}}$. The expected length of CI_s is therefore

$$\begin{aligned} E(CI_s) &= E\left(2d\sqrt{\frac{nS^2}{(n+\tau^2)^2}}\right) \\ &= 2d\sqrt{\frac{n}{(n+\tau^2)^2}}E(S) \\ &= 2d\sqrt{\frac{n}{(n+\tau^2)^2}}\frac{\sqrt{n-1}\Gamma((n-1)/2)}{\sqrt{2}\Gamma(n/2)}\sigma \end{aligned}$$

4) Coverage probability and Expected length of CI_p: 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 by

$$\begin{aligned} & P(\mu \in CI_p) \\ = & P\left[\frac{1}{\tau}\sqrt{\frac{(n-1)S^2}{u}} < \mu < \frac{1}{\tau}\sqrt{\frac{(n-1)S^2}{v}}\right] \\ = & P\left[-\frac{1}{\tau}\sqrt{\frac{(n-1)S^2}{v}} < -\mu < -\frac{1}{\tau}\sqrt{\frac{(n-1)S^2}{u}}\right] \\ = & P\left[\bar{X} - \frac{1}{\tau}\sqrt{\frac{(n-1)S^2}{v}} < \bar{X} - \mu < \bar{X} - \frac{1}{\tau}\sqrt{\frac{(n-1)S^2}{u}}\right] \\ = & P\left[\frac{\bar{X}}{Q} - \frac{1}{\tau Q}\sqrt{\frac{(n-1)S^2}{v}} < \frac{\bar{X} - \mu}{Q} < \right. \\ & \left. \frac{\bar{X}}{Q} - \frac{1}{\tau Q}\sqrt{\frac{(n-1)S^2}{u}}\right] \\ = & P\left[\frac{\bar{X}}{Q} - \frac{1}{\tau Q}\sqrt{\frac{(n-1)S^2}{v}} < Z < \right. \\ & \left. \frac{\bar{X}}{Q} - \frac{1}{\tau Q}\sqrt{\frac{(n-1)S^2}{u}}\right] \end{aligned}$$

$$
\begin{aligned}
&= P[C_1 < Z < C_2] \\
&= E[I_{\{C_1<Z<C_2\}}(\tau)] \\
&= E[E[I_{\{C_1<Z<C_2\}}(\tau)]|S^2] \\
&= E[\Phi(C_2) - \Phi(C_1)]
\end{aligned}
$$

where

$$
I_{\{C_1<Z<C_2\}}(\tau) = \begin{cases} 1, & \text{if } \tau \in \{C_1 < Z < C_2\} \\ 0, & \text{otherwise} \end{cases}
$$

$C_1 = \frac{\bar{X}}{Q} - \frac{1}{\tau Q}\sqrt{\frac{(n-1)S^2}{v}}$, $C_2 = \frac{\bar{X}}{Q} - \frac{1}{\tau Q}\sqrt{\frac{(n-1)S^2}{u}}$, $Q = \sigma/\sqrt{n}$.

The expected length of CI_p is therefore

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

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 Monte Carlo simulation to assess three confidence intervals notified in the previous section: CI_μ, CI_s and CI_p based on their average length widths. We design a simulation, without losing generality, by setting $\sigma = 1$, $\tau = 0, 0.01, 0.05, 0.5, 0.1, 0.3, 0.5, 0.8, 0.9, 1, 1.1, 1.3, 1.5, 1.9, 2.5$ and the samples sizes $n = 10, 30, 50, 100$. We wrote function in R program to generate the data which is normally distributed with means and variances which are mentioned previously to construct three confidence intervals, i.e. CI_μ, CI_s and CI_p and then compute average length width of each confidence interval. All results are illustrated in Table I.

VI. Simulation results

From Table I, we found that the expected length the CI_μ when known coefficient of variation is wider than those two new confidence intervals for every sample sizes. For known coefficient of variation (τ) and small sample sizes i.e. $n = 10, 30$, the confidence interval CI_s is shorter than the confidence interval CI_p when $\tau \leq 0.5$, otherwise the confidence interval CI_p is preferable. For $n = 50, 100$, the confidence interval CI_p is shorter than that confidence interval CI_s when $\tau \geq 0.8$.

TABLE I
THE EXPECTED LENGTHS OF CI_μ, CI_s AND CI_p

τ	$E(CI_\mu)$	$E(CI_s)$	$E(CI_p)$	$E(CI_\mu)$	$E(CI_s)$	$E(CI_p)$
		$n=10$			$n=30$	
0.00	1.4709	1.2744	-	0.7532	0.7218	-
0.01	1.4709	1.2744	116.975	0.7532	0.7218	55.2651
0.05	1.4709	1.2741	23.3951	0.7532	0.7218	11.0530
0.10	1.4709	1.2731	116975	0.7532	0.7216	5.5265
0.30	1.4709	1.2630	3.8991	0.7532	0.7197	1.8421
0.50	1.4709	1.2433	2.3395	0.7532	0.7159	1.1053
0.80	1.4709	1.1977	1.4621	0.7532	0.7067	0.6908
0.90	1.4709	1.1789	1.2997	0.7532	0.7028	0.6140
1.00	1.4709	1.1585	1.1697	0.7532	0.6985	0.5526
1.10	1.4709	1.1368	1.0634	0.7532	0.6938	0.5024
1.30	1.4709	1.0901	0.8998	0.7532	0.6833	0.4251
1.50	1.4709	1.0403	0.7798	0.7532	0.6715	0.3684
1.90	1.4709	0.9363	0.6156	0.7532	0.6443	0.2908
2.50	1.4709	0.7842	0.4679	0.7532	0.5974	0.2210
		$n=50$			$n=100$	
0.00	0.5713	0.5571	-	0.3978	0.3978	-
0.01	0.5713	0.5571	41.2900	0.3978	0.3929	28.7385
0.05	0.5713	0.5571	8.2580	0.3978	0.3929	5.6877
0.10	0.5713	0.5570	4.1290	0.3978	0.3929	2.8438
0.30	0.5713	0.5561	1.3763	0.3978	0.3926	0.9479
0.50	0.5713	0.5544	0.8258	0.3978	0.3920	0.5687
0.80	0.5713	0.5501	0.5161	0.3978	0.3904	0.3554
0.90	0.5713	0.5483	0.4587	0.3978	0.3898	0.3159
1.00	0.5713	0.5462	0.4190	0.3978	0.3890	0.2843
1.10	0.5713	0.5440	0.3753	0.3978	0.3882	0.2585
1.30	0.5713	0.5389	0.3189	0.3978	0.3864	0.2187
1.50	0.5713	0.5332	0.2752	0.3978	0.3843	0.1895
1.90	0.5713	0.5196	0.2173	0.3978	0.3792	0.1496
2.50	0.5713	0.4952	0.1651	0.3978	0.3698	0.1137

VII. Conclusion

In this paper we proposed new confidence intervals for the normal population mean with known coefficient of variation. We derived, mathematically, coverage probabilities and expected lengths of these intervals. It is shown in sections IV that the coverage probabilities of CI_μ and CI_s and CI_p are equal to the nominal value $1-\alpha$. To compare confidence intervals between CI_μ and CI_s and CI_p, one can compare the expected lengths of each interval derived in sections IV. It is recommended that for known small coefficient of variation i.e. $\tau \leq 0.5$, the confidence interval CI_s is preferable otherwise we choose the confidence interval CI_p. Further research may consider statistical estimation for the difference between normal means and normal variances with known coefficients of variation. Also one might consider to derive their coverage probabilities and expected lengths as shown in this paper.

Acknowledgment

The author is grateful to King Mongkut's University of Technology North Bangkok for their partial 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] D. T. Searls, A Note on the Use of an Approximately Known Coefficient of Variation, The American Statistician, 21(1967), 20-21.
[6] 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.
[7] R.E. Walpole, R. H. Myers, S.L. Myers, K. Ye., Probability & Statistics for Engineers & Scientists, Prentice Hall, New Jersey, 2002.