

An examination of the variability of agricultural production using a new approximate confidence interval for the coefficient of variation

Vorrapot Saelee^{a,*}, Suchada Kornpetpanee^a, Patcharee Wongkasem^b

^a Research and Statistics in Cognitive Science, Burapha University, 169 Longhad Road, Bangsaen, Chonburi 20131 Thailand

^b Faculty of Science, Burapha University, Thailand

*Corresponding author, e-mail: vorrapot@gmail.com, vorrapot@windowslive.com

Received 14 Mar 2012

Accepted 1 Feb 2013

ABSTRACT: The objective of this study was to develop a new approximation method for determining confidence intervals for the coefficient of variation, and, using Monte Carlo simulation, to compare the relative efficiency of the new method with Miller's, Vangel's, and two types of Mahmoudvand and Hassani's methods. It was found that the efficiency of the new approximation method was superior when $n \geq 30$ and the population coefficient was greater than 0.5. With $n \geq 30$ and the population coefficient less than 0.5, the new method was no less efficient than the other methods, and sometimes better. The new method was applied to examine the variability of a variety of agricultural products using data from 2006–2010. It was found that the variation in cassava yields was less than that for maize, rubber, and main rice. Cassava production was more stable across all environments.

KEYWORDS: new approximation method, normal distribution, Monte Carlo Technique, asymptotically consistent

INTRODUCTION

The coefficient of variation is an important measurement of reliability of data¹. It is defined by $\kappa = \sigma/\mu$ where σ is the population standard deviation and μ is the population mean. Given the random variable X is normally distributed, the represented sample estimate of κ , depends on the sample size n . The statistic is $\hat{\kappa} = S/\bar{x}$, where S is the standard deviation, and \bar{x} is the sample mean. The sample coefficient of variation, $\hat{\kappa}$ is measured as a point estimate of κ . In some cases, using the coefficient of variation may provide a poor estimate of κ . Hence the researcher should use it carefully. One way is to use the confidence intervals for the coefficient of variation by considering the width of the range that is indicative of an environment suitable for agricultural production, such as the variability of agricultural production by the confidence intervals for coefficient of variation². This article examines and compares different methods. The new approximation method is compared with Miller's method, Vangel's method and two types of Mahmoudvand and Hassani's method by Monte Carlo simulation. It also examines the variability of agricultural products in different regions using the new approximate confidence intervals for the coefficient of variation.

Table 1 Crop yield statistics (values are in kg/km²).

Crop type	Max	Min	Mean	Standard deviation
cassava	2 575 000	1 573 750	2 067 056.3	192 243.8
rubber	210 000	69 375	154 806.3	24 300.0
maize	451 250	274 375	393 006.3	42 075.0
main rice	560 000	181 875	296 425	86 843.8

DESCRIPTION OF THE DATA

Data sets from agricultural products in 77 provinces collected from 2006–2010 in different regions in Thailand were used³. The data sets were from cassava, rubber, maize and main rice. The numbers of data points were 226 for cassava, 282 for rubber, 200 for maize, and 380 for main rice (Table 1).

METHOD

Previous approximation of confidence intervals for coefficient of variation

There have been several approximation methods. Miller^{4,5} proposed the following:

$$CI1 = \frac{s}{\bar{x}} \pm Z_{\alpha/2} \left(\frac{s}{\bar{x}} \right) \nu^{-1/2} \left[0.5 + \left(\frac{s}{\bar{x}} \right)^2 \right]^{1/2} \quad (1)$$

where $\nu = n - 1$, and Z_{α} is the statistic of the standardized normal distribution corresponding to $(1 -$

α) 100%. Vangel's coefficient of variation is given by

$$CI2 = \hat{\kappa} \left[\left(\frac{u+2}{\nu+1} - 1 \right) \hat{\kappa}^2 + \frac{u}{\nu} \right]^{-1/2} \quad (2)$$

where $u = u_1, u_2$ and $u_1 = \chi_{\nu, 1-\alpha/2}^2$ and $u_2 = \chi_{\nu, \alpha/2}^2$. Mahmoudvand and Hassani⁷ suggested two new confidence intervals for the coefficient of variation in a normal distribution:

$$CI4 = \frac{\hat{\kappa}}{2 - c_n \pm Z_{1-\alpha/2}(1 - c_n^2)^{1/2}} \quad (3)$$

$$CI5 = \hat{\kappa} \pm \frac{\hat{\kappa}}{2 - c_n} Z_{1-\alpha/2} \left[(1 - c_n^2) + \frac{\hat{\kappa}^2}{n} \right]^{1/2} \quad (4)$$

Miller's formula is usable when $n \leq 10$ and $0.33 \leq \kappa \leq 0.67$. The Mahmoudvand and Hassani methods give results to Miller's method for large n . This paper proposes a new confidence interval CI3.

The new of approximation confidence intervals for coefficient of variation

In this section, we shall consider the mean and variance of the coefficient of variation for normal distribution. We introduce an asymptotically unbiased estimator for κ , obtain its variance and finally construct a confidence interval. X is a normal random variable with mean μ and variance σ^2 . $E(X) = \mu$, $V(X) = \sigma^2$. For a random sample of size n , $E(\bar{x}) = \mu$ and $V(\bar{x}) = (\sigma^2/n)$. The sample variance is $S^2 = \sum_{i=1}^n ((X_i - \bar{X})^2 / (n - 1))$. From Ref. 10, $\hat{\kappa} = S_n / \bar{x}_n$. Hence $E(\hat{\kappa}) = E(\sum_{i=1}^n \hat{\kappa})$. We now show that $E(\hat{\kappa}) = \kappa$.

$$E\left(\frac{x}{y}\right) = \left(\frac{E[x]}{E[y]}\right) \left(1 + \frac{\text{Var}(y)}{(E(y))^2}\right).$$

$$E(\hat{\kappa}) = E\left[\frac{s}{\bar{x}}\right] = \left(\frac{E[s]}{E[\bar{x}]}\right) \left(1 + \frac{\text{Var}(\bar{x})}{(E(\bar{x}))^2}\right).$$

$E(S) = C_n(n) \sigma$. From Refs. 7, 9

$$C_n = \sqrt{\frac{2}{n-1} \frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})}}$$

$$= 1 - \frac{1}{4n} - \frac{7}{31n^2} - \frac{19}{128n^3} + O(n^{-4}).$$

$$E(\hat{\kappa}) = \kappa \left(1 - \frac{1}{4n} - \frac{7}{31n^2} - \frac{19}{128n^3}\right) \left(1 + \frac{\kappa^2}{n}\right)$$

As $n \rightarrow \infty$, $E(\hat{\kappa}) \rightarrow \kappa$.

$$E(S) = C_n(n)\sigma = \left(1 - \frac{1}{4n} - \frac{7}{32n^2} - \frac{19}{128n^3}\right)\sigma$$

and $E(S^2) = \sigma^2$. $V(S) = E(S^2) - [E(S)]^2 = Q\sigma^2$ where

$$Q = \frac{1}{2n} + \frac{3}{8n^2} + \frac{3}{16n^3} - \frac{125}{1024n^4} - \frac{133}{2048n^5} - \frac{361}{16384n^6}.$$

Hence⁸

$$V\left(\frac{x}{y}\right) = \left(\frac{E(x)}{E(y)}\right)^2 \left(\frac{V(x)}{[E(x)]^2} + \frac{V(y)}{[E(y)]^2}\right)$$

$$V\left(\frac{S}{\bar{X}}\right) = \left(\frac{E(S)}{E(\bar{X})}\right)^2 \left(\frac{V(S)}{[E(S)]^2} + \frac{V(\bar{X})}{[E(\bar{X})]^2}\right).$$

Thus

$$V\left(\frac{S}{\bar{x}}\right) = \kappa^2 Q + (1 - Q) \left(\frac{\kappa^4}{n}\right).$$

Hence

$$V\left(\frac{S}{\bar{X}}\right) = \left(\frac{4n+3}{8n^2}\right)\kappa^2 + \left(\frac{8n^2-4n-3}{8n^3}\right)\kappa^4.$$

$X \sim N(\mu, \sigma^2)$ and $\bar{x} \sim N(\mu, (\sigma^2/n))$. Therefore $\bar{\kappa}$ is also asymptotically consistent. Now, using the normal approximation, we have

$$Z = \frac{\hat{\kappa} - \kappa}{[\text{Var}(\hat{\kappa})]^{1/2}} \rightarrow N(0, 1), \quad (5)$$

where

$$\text{Var}(\hat{\kappa}) = \left(\frac{4n+3}{8n^2}\right)\kappa^2 + \left(\frac{8n^2-4n-3}{8n^3}\right)\kappa^4$$

Hence the 100(1 - α)% confidence interval for κ based on (5) is

$$\hat{\kappa} - Z_{\alpha/2}[\text{Var}(\hat{\kappa})]^{1/2} < \kappa < \hat{\kappa} + Z_{\alpha/2}[\text{Var}(\hat{\kappa})]^{1/2}$$

If κ is unknown, we use $\hat{\kappa}$ instead. Hence the 100(1 - α)% confidence interval for κ is

$$\hat{\kappa} - Z_{\alpha/2}[\text{Cov}(\hat{\kappa})]^{1/2} < \kappa < \hat{\kappa} + Z_{\alpha/2}[\text{Cov}(\hat{\kappa})]^{1/2} \quad (6)$$

where

$$\text{Cov}(\hat{\kappa}) = \left(\frac{4n+3}{8n^2}\right)\hat{\kappa}^2 + \left(\frac{8n^2-4n-3}{8n^3}\right)\hat{\kappa}^4$$

or

$$CI3 = \hat{\kappa} \pm Z_{\alpha/2}[\text{Cov}(\hat{\kappa})]^{1/2}. \quad (7)$$

RESULTS AND DISCUSSION

Data was simulated with a Monte Carlo simulation. The simulation was repeated 50 000 times. The sample size was 10, 15, 25, 30, 50, and 100. The coefficient of variation of population was 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9 at confidence level of 90% and 95%.

Table 2 Average width and coverage probability (in brackets) of the new method for $1 - \alpha = 0.90, 0.95$ at large sample size ($n \geq 30$) and $\kappa = 0.1-0.9$.

n	30	50	100	500
90%				
0.1	0.047(0.906)	0.033(0.893)	0.024(0.898)	0.011(0.901)
0.2	0.089(0.902)	0.069(0.893)	0.048(0.898)	0.022(0.901)
0.3	0.139(0.900)	0.108(0.894)	0.076(0.899)	0.034(0.902)
0.4	0.197(0.896)	0.152(0.895)	0.107(0.900)	0.048(0.901)
0.5	0.262(0.898)	0.204(0.895)	0.143(0.901)	0.064(0.901)
0.6	0.337(0.896)	0.262(0.900)	0.185(0.901)	0.082(0.902)
0.7	0.420(0.895)	0.328(0.907)	0.232(0.901)	0.103(0.901)
0.8	0.513(0.896)	0.402(0.905)	0.285(0.902)	0.126(0.900)
0.9	0.614(0.901)	0.483(0.905)	0.346(0.902)	0.152(0.900)
95%				
0.1	0.051(0.952)	0.040(0.942)	0.028(0.946)	0.013(0.950)
0.2	0.106(0.951)	0.082(0.952)	0.058(0.946)	0.026(0.950)
0.3	0.166(0.941)	0.128(0.953)	0.091(0.946)	0.040(0.950)
0.4	0.235(0.940)	0.181(0.955)	0.128(0.946)	0.057(0.950)
0.5	0.316(0.949)	0.241(0.955)	0.171(0.946)	0.076(0.949)
0.6	0.409(0.947)	0.309(0.952)	0.220(0.945)	0.098(0.949)
0.7	0.497(0.949)	0.387(0.951)	0.276(0.945)	0.122(0.950)
0.8	0.608(0.953)	0.474(0.948)	0.340(0.944)	0.150(0.950)
0.9	0.695(0.951)	0.569(0.953)	0.412(0.944)	0.181(0.950)

Table 3 Average width and coverage probability (in brackets) of the new method for $1 - \alpha = 0.90, 0.95$ at a small sample size ($n < 30$) and $\kappa = 0.1-0.9$.

n	10	15	25
90%			
0.1	0.075(0.850)	0.061(0.871)	0.047(0.882)
0.2	0.155(0.851)	0.126(0.872)	0.097(0.882)
0.3	0.244(0.852)	0.199(0.889)	0.153(0.884)
0.4	0.346(0.851)	0.282(0.883)	0.217(0.884)
0.5	0.441(0.882)	0.379(0.893)	0.291(0.896)
0.6	0.506(0.899)	0.476(0.896)	0.378(0.897)
0.7	0.590(0.900)	0.568(0.901)	0.458(0.897)
0.8	0.638(0.903)	0.631(0.900)	0.558(0.905)
0.9	0.684(0.905)	0.682(0.899)	0.650(0.909)
95%			
0.1	0.089(0.910)	0.073(0.922)	0.056(0.937)
0.2	0.184(0.920)	0.150(0.932)	0.116(0.936)
0.3	0.291(0.928)	0.237(0.941)	0.237(0.941)
0.4	0.404(0.945)	0.336(0.948)	0.258(0.938)
0.5	0.489(0.943)	0.446(0.946)	0.347(0.943)
0.6	0.574(0.952)	0.539(0.949)	0.449(0.941)
0.7	0.621(0.954)	0.625(0.950)	0.543(0.953)
0.8	0.664(0.955)	0.677(0.949)	0.649(0.953)
0.9	0.707(0.944)	0.722(0.951)	0.713(0.949)

Comparison of the methods from simulation

The results of using Monte Carlo simulations to investigate the coverage probabilities of the confidence intervals CI1, CI2, CI3, CI4 and CI5 and their average width of confidence intervals are presented in Table 2 and Table 3.

The new method of confidence intervals for coefficient of variation was derived from the theory of large samples. As a result of comparing the performance of a new method with Miller’s method, Vangel’s method and Mahmoudvand and Hassani’s method in the sample sizes of 30, 50 and 100, the new

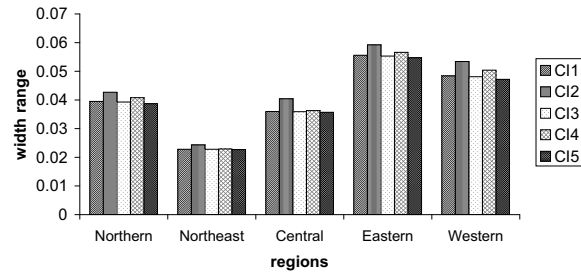


Fig. 1 Width range of Miller’s method (CI1), Vangel’s method(CI2), new method (CI3) and two types of Mahmoudvand and Hassani’s method (CI4 and CI5) at 95% confidence level by regions of cassava.

method in most cases is more efficient than the other methods, The new method has a coverage probability which is closer to the nominal level than the other methods and the new method has an average width less than the other methods. Therefore we concluded that the new method in most cases was more efficient than Miller’s method, Vangel’s method and Mahmoudvand and Hassani’s methods in two ways. First, the sample size has less than 30 and population coefficient of variation is 0.5–0.9. Second, the sample size $n \geq 30$ and population coefficient of variation is 0.1–0.9. The new method shows better results when there is a larger sample size. The new method has coverage probability close to nominal level when the sample size is large. When the sample size is less than 30 the new method has coverage probability close to the nominal level when the population coefficient of variation is between 0.5 and 0.9, as shown in Table 2 and Table 3.

Comparison of approximation methods of data from agricultural yield

At 95% confidence levels for cassava yield, the north-east region has the smallest width range of all methods which means that it is the most stable across environments in Thailand for cassava. If the width ranges of each method were compared, all methods have almost the same width ranges (Fig. 1).

At 95% confidence levels for rubber yield, the southern region has smallest width range of all methods, which means that it is the most stable across environments in Thailand for rubber (Fig. 2).

At 95% confidence levels for maize yield, the eastern region has the smallest width range of all methods, which means that it is the most stable across environments in Thailand for maize (Fig. 3).

At 95% confidence levels for main rice yield, the northeast region has the smallest width range of all

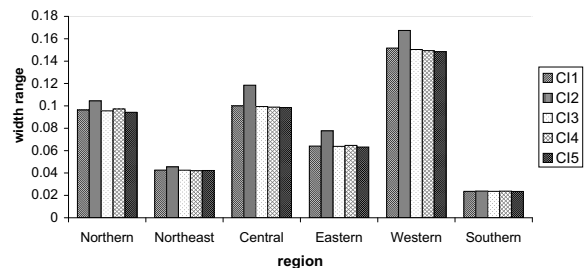


Fig. 2 Width ranges at 95% confidence level by regions of rubber.

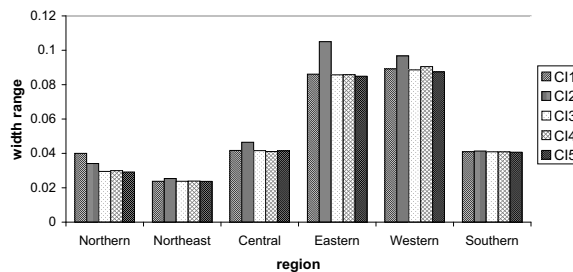


Fig. 4 Width ranges at 95% confidence level by regions of main rice.

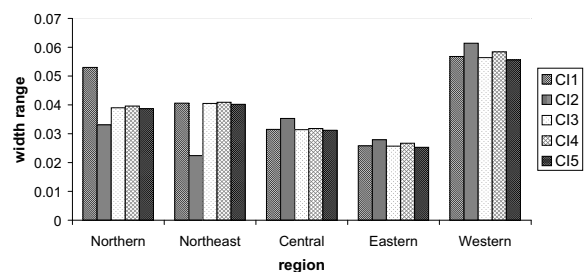


Fig. 3 Width ranges at 95% confidence level by regions of maize.

Table 4 Variability of agricultural product from 2006–2010 using by the new approximation confidence intervals for coefficient of variation at 95% confidence level.

Crop type	95% confidence level		
	<i>L</i>	<i>U</i>	<i>R</i>
cassava	0.0843	0.1017	0.0173
rubber	0.1437	0.1703	0.0266
maize	0.0964	0.1176	0.0212
main rice	0.2702	0.3153	0.0451

methods, which means that it is the most stable across environments in Thailand for main rice (Fig. 4).

L is the lower limit, *U* is the upper limit, $R = U - L$. In Table 4, when the approximation confidence interval for coefficient of variation was used, the examination the variable of agricultural products found that cassava was more stable across environments than maize, rubber, and main rice.

Acknowledgements: The authors thank the Office of the Higher Education Commission, which provided scholarships at the doctoral level for the main authors. The author thanks the editor and the referees for their constructive comments. We thank Suchada Kornpetanee for insightful suggestions.

REFERENCES

- Gomez KA, Gomez AA (1984) *Statistical Procedures for Agricultural Research*, 2nd edn, Wiley, New York.
- Taye G, Njuho P (2008) Monitoring field variability using confidence interval for coefficient of variation. *Comm Stat Theor Meth* **37**, 831–46.
- Office of Agricultural Economics Ministry of Agriculture Cooperatives (2010) [www.oae.go.th/main.php?filename=agri_production].
- Miller EG (1991) Asymptotic test statistics for coefficient of variation. *Comm Stat Theor Meth* **20**, 3351–63.
- Ng CK (2005) *Performance of three methods of interval estimation of the coefficient of variation*.

Department of Management Sciences City, Univ of Hong Kong [interstat.statjournals.net/YEAR/2006/articles/0609002.pdf].

- Vangel MG (1996) Confidence intervals for a normal coefficient of variation. *Am Statistician* **50**, 21–6.
- Mahmoudvand R, Hassani H (2009) Two new confidence intervals for the coefficient of variation in a normal distribution. *J Appl Stat* **36**, 429–42.
- Blumenfeld D (2001) *Operations Research Calculations Handbook*, Boca Raton, New York.
- Engineering Statistics Handbook (2003) [www.itl.nist.gov/div898handbook/eda/eda.d.htm].
- McKay AT (1932) Distribution of the coefficient of variation and the extended “t” distribution. *J Roy Stat Soc Stat Soc* **95**, 695–8.