Available online at http://scik.org

Commun. Math. Biol. Neurosci. 2021, 2021:65

https://doi.org/10.28919/cmbn/5917

ISSN: 2052-2541

EFFICIENCY PROPORTION ESTIMATIONS FOR THAI PEOPLE INFECTED CORONAVIRUS 2019 (COVID-19) IN THAILAND BY JACKKNIFING METHOD AND BOOTSTRAPPING METHOD

THAMMARAT PANITYAKUL[1], RONNASON CHINRAM[1], WANDEE WANISHSAKPONG[2,*]

[1]Division of Computational Science, Faculty of Science, Prince of Songkla University,

Hat Yai, Songkhla 90110, Thailand

[2]Department of Statistics, Faculty of Science, Kasetsart University, Bangkok 10900, Thailand

Abstract: This research aims to estimate the proportion of Thai people infected with COVID-19 and to compare the efficiency of estimates between the Jackknifing method and the Bootstrapping method. Data sample sizes of 100, 200, 400, 800 and 1,609 were obtained from the Department of Disease Control in March 2020, then repeatedly resampled by each method for 1,000 times. The proportion estimated of Thai people infected with COVID-19 by the Jackknifing method was 91.05034% with the width of the confidence interval at 0.06219% percent and the coefficient of variation of 0.0195%. The estimate of the proportion infected using the Bootstrapping method was 91.04842% with the width of the confidence interval 2.86047% and the coefficient of variation at 0.78527%. It was found that the estimations of proportion by the Jackknifing Method had a lower coefficient of variation and a smaller width of the confidence interval than did the Bootstrapping Method and increasing sample sizes led to a decrease in the coefficient of variation and the width of the confidence interval. The results found that the estimate of proportion infected was more efficient with the Jackknifing Method than with the Bootstrapping Method.

*Corresponding author

E-mail address: wandee.w@ku.ac.th

Received April 23, 2021

Keywords: COVID-19; estimation of proportion; jackknifing method; bootstrapping method.

2010 AMS Subject Classification: 62F40.

1. INTRODUCTION

Currently all over the world there is a focus on contagious rogue viruses caused by the most recently discovered virus known as the Coronavirus, COVID-19. The disease can be transmitted from person to person through the nostrils. As of now, the COVID-19 has a widespread outbreak across the Center for Emergency Operations, the Department of Disease Control (2020), with a worldwide list of confirmed cases totaling 22,069,384 with 777,751 deaths as of March 2020. In Thailand there were 3,381 confirmed cases of COVID-19with 58 deaths of people having a median age of 36 years. By gender, (1 month-97 years) males comprised 1,892 cases and females 1,489 cases. There were 3,032 cases for people of Thai nationality and 340 for other nationalities with 9 of unknown nationality. Of these cases there were 208 cases of personal disease [I have no idea what this means] and 3,173 non-medical histories [does this mean no underlying pre-existing medical conditions or with those conditions??], history of common diseases such as hypertension, allergies, diabetes, etc.

According to the data, the patients were asymptomatic and the number of cases increased steadily from January, 2020. Both Thai and foreign patients increased in number costing the state very severe treatment costs, not including the loss of the quality of life of people in the general public forced into quarantine or who lost their income. In addition, the data have not yet specified the ratio of Thai to foreigners infected with COVID-19 or the proportion of disease incidence in each region of Thailand. Due to the statistical report of the high number of people infected with the COVID-19 in March 2020 the researchers were interested in estimating the proportion of Thais infected with the COVID-19 in Thailand in that month. Juthamas and others (2014) studied the method for estimating the proportion of hypertensive disease incidence among people living in Doi Tao District, Chiang Mai Province and compared the efficiency of estimation of proportions between the Bootstrap method and Jackknife method. The findings were that the proportion estimation using the Jackknife method had an efficiency of variation lower than the bootstrap

method which was consistent with other studies. Therefore, this research is interested in estimating the proportion of Thai people infected with the COVID-19 in Thailand and comparing the efficiency of the estimation of the proportion between the Jackknife method and the Bootstrap method as a guide for selecting methods for estimating the proportion of Thai people infected with COVID-19 in Thailand.

2. PRELIMINARIES

In this study, the population was people diagnosed with COVID-19 living in Thailand. The population size was 1,609 people. The defined dependent variable (Y) where Y = 0 was Thai nationals diagnosed with COVID-19 in Thailand, Y = 1 was foreigners diagnosed with COVID-19 in Thailand, and independent variables were gender (X_1) (X_1 = 0 is male, X_1 = 1 is female), age (X_2) (X_2 = 1 was under 18 years, X_2 = 2 was 18 - 34 years, X_2 = 3 was 35 - 60 years, X_2 = 4 was 60 over years old) and regions (X_3) (X_3 = 1 was the North, X_3 = 2 was the Northeast, X_3 = 3 was the Central region, X_3 = 4 was the East, X_3 = 5 was the West, X_3 = 6 was the South). The data were obtained from the Emergency Operation Center of the
Department of Disease Control using data for March 2020. Random sample sizes were used from specifying the sample size to 100 in increments of 100 Thus the sample sizes were 100, 200, 400, 800, and the size equal to the total population size was 1,609 people, setting a random round to be repeated 1,000 times by defining the degree of significance (α) 0.05. Data analysis was conducted using the R program and the Excel program.

Basic descriptive statistics such as frequency, percentage of dependent, and independent variables to be studied were analyzed. Simple, non-return sampling at sizes 100, 200, 400, 800 and equal to the total population size of 1,609 people and point and range estimates were performed by the Jackknife and Bootstrap methods. In comparison estimation was determined from the request coefficient variation (C.V.) and the amplitude of the range c. The confidence (length) of both methods in each sample size as determined by the Bootstrap method determined the number of rounds that were randomly repeated 1000 times and concluded the comparison. Estimates of the

proportion (p) of Thai people diagnosed with COVID-19 in Thailand were classified by the independent variables of sex, age, and region, and summarized the proportion estimate.

The Jackknife method was proposed by Quenouille (1956) to reduce the bias of the estimator, which was able to find the approximation of the parameters and variance estimates that the Jackknife method had. Collected and developed further by Miler (1974), the Jackknife method is a re-sampling method that selects new samples from a single random sample. Take one sample, x_1, x_2,…, x_n from a population of different distributions that have the density function f and the distribution function of F. Let θ be the approximately required parameter in this population and let $\hat{\theta}_j$ be the value. Estimation of parameter with jackknife method calculated from sample data of size n. Finding the Jackknife Method Estimation Let's start by eliminating the x_1 value from the sample. Then calculate the approximate value of θ from x_2, x_3,…, x_n, you will get an approximation of $\hat{\theta}_{(1)}$ and cut the x_2 from the sample, then calculate the approximate value of θ from x_1, x_3,…,x_n will get an approximation of $\hat{\theta}_{(2)}$ Calculate the approximation of θ until all n times, the approximate value of θ for n numbers is $\hat{\theta}_{(1)}, \hat{\theta}_{(2)}, \ldots, \hat{\theta}_{(n)}$.

Let $\hat{\theta}_J$ be an approximation of the parameter θ with the jackknife method, where point approximation is determined by

$$\hat{\theta}_J = \frac{\sum_{i=1}^{n} \hat{\theta}_{(i)}}{n}$$

Estimation of the range parameter θ by jackknife method at significant scale or 100% convinced (1– α) gives that

$$P\left(\hat{\theta}_{JL} < \theta < \hat{\theta}_{JU}\right) = 1 - \alpha$$

where $\hat{\theta}_{JL}$ is the lower confidence limit, $\hat{\theta}_{JU}$ is the upper confidence limit which is calculated from the approximate value of $\hat{\theta}_{(1)}, \hat{\theta}_{(2)}, \ldots, \hat{\theta}_{(n)}$ that has been arranged in ascending order and then calculated at the 100th percentile position (α / 2), is defined as $\hat{\theta}_{JL}$ and is determined at the 100th percentile position (1-α/2), set as $\hat{\theta}_{JU}$. So a 100% confidence interval (1– α) is achieved with the

jackknife method is $[\hat{\theta}_{JL}, \hat{\theta}_{JU}]$.

The bootstrap method is a method of estimating parameters using repeated sampling from a single set of samples as well as the method of jackknife, but the bootstrap re-sampling method is used to generate a new set of samples from a single batch sample using n-size reconstituted sampling. This method was proposed by Efron (1979) and was later developed by Efron (1982). A single random sample of size n containing generated a possible set of n-size samples, that is, instead of sampling repeatedly from the population, using sampling the data by doing the following methods. The n samples, is $x_1, x_2, \ldots, x_n$ are independent from populations with different distributions. Let θ be the approximate desired parameter in this population, and $\hat{\theta}_B$ is the approximate value of parameter θ. By using the bootstrap method calculated from sample data of size n, creating a function enumerate by giving the probability of x_i , i = 1,2, …, n is $\frac{1}{n}$. This is called the empirical distribution function. Perform randomization of the returning n-size sample from the resulting empirical distribution function, that is a reconstituted n-size sampling was performed from the resulting empirical distribution function, that is, the sampling was sampled one number at a time, n times from the sample set $x_1, x_2, \ldots, x_n$. The values are returned to the sample set prior to the next sampling. A set of values of x_i, i = 1,2,…, n occurred more than $x_1^*, x_2^*, \ldots, x_n^*$. It is a set of sample sizes n, which called the bootstrap example. Repeat the same procedure, each time getting an example. Estimation using the bootstrap method be started by sampling 1 value at a time returned at n times from the sample set, $x_1^*, x_2^*, \ldots, x_n^*$ and calculate the approximate value of θ to get an approximate value of $\hat{\theta}_1^*$. Then take sampling one by one 1 Returned value at n times from the sample set gives $x_1^*, x_2^*, \ldots, x_n^*$. And calculate an approximation of θ to get an approximation of $\hat{\theta}_2^*$. Calculate and estimate θ until B times, will get an approximation of the B numbers is $\hat{\theta}_1^*, \hat{\theta}_2^*, \ldots, \hat{\theta}_B^*$. And then create a histogram, assigning each one to have equal probability 1B. The distribution of the bootstrap sample statistic is obtained. Let $\hat{\theta}_B$ be an approximation of the θ parameter with the bootstrap method. In which point estimation is determined by

$$\hat{\theta}_B = \frac{\sum_{i=1}^{B} \hat{\theta}_i^*}{B}$$

A range approximation of the parameter θ by bootstrap method at the significance scale α or at the confidence interval (1-α) 100% gives that

$$P\left(\hat{\theta}_{BL} < \theta < \hat{\theta}_{BU}\right) = 1 - \alpha$$

where $\hat{\theta}_{BL}$ is the lower confidence limit, $\hat{\theta}_{BU}$ is the upper confidence limit.

This is obtained from the distribution of the bootstrap example $\hat{\theta}_1^*, \hat{\theta}_2^*, \ldots, \hat{\theta}_B^*$, which arranged from the smallest value to find the larger and then calculate the value at the 100th percentile position (α/2), set to $\hat{\theta}_{BL}$ and find the value at the 100th percentile position (1 - α/2), set it to $\hat{\theta}_{BU}$, therefore, a 100% confidence interval (1-α) is obtained with the bootstrap method, $[\hat{\theta}_{BL}, \hat{\theta}_{BU}]$

Comparison of point estimation methods based on the coefficient of variation or CV(Jutamas et al, 2014). The lower coefficient of variation was considered more efficient. Coefficient of variation Variables which can be calculated by the coefficient of the variation of the Jackknife method.

$$\text{C.V.} = \frac{\text{S.D.}}{\bar{X}}$$

where S.D. = $\sqrt{\frac{\sum_{i=1}^{n}\left(\hat{\theta}_{(i)} - \hat{\theta}_J\right)^2}{n-1}}$ and $\bar{X} = \frac{\sum_{i=1}^{B} \hat{\theta}_{(i)}}{n}$ for the Jackknife method, and S.D. = $\sqrt{\frac{\sum_{i=1}^{B}\left(\theta_i^* - \hat{\theta}_B\right)^2}{B-1}}$ and $\bar{X} = \frac{\sum_{i=1}^{B} \hat{\theta}^*(i)}{B}$ for the bootstrap method.

For performance comparison, interval estimation will determined by the width of the confidence interval. It is denoted by the length calculated from the difference between the upper bound and the lower bound of the confidence interval. And the method that yields the narrowest confidence interval is considered to be a more appropriate estimation of the confidence interval. The width of the confidence interval of the jackknife Method:

$$\text{Length} = \hat{\theta}_{JU} - \hat{\theta}_{JL}$$

The width of the confidence interval of the bootstrap method:

$$\text{Length} = \hat{\theta}_{BU} - \hat{\theta}_{BL}$$

3. MAIN RESULTS

As general information, there were 1,609 people, 940 males, and 669 females, with 36 under age 18, 689 in the 18 -34 years range, 715 people in the 35 -60 years range and 169 aged 60 and over. By region of Thailand, there were 49 people from the North, 84 people from the Northeast, 1177 from the Central Region, 72 people from the East, 27 people from the West and 200 people from the South 200.

A comparison of the performance of the Jackknife method and the bootstrap Method can be made from the width of the confidence interval (Length). The results are shown in Table1, Figure 1 and Figure 2 as follows.

Table 1 A Comparison of Estimations of Proportion (COVID-19) for Thai in Thailand by C.V and length

Estimate of Proportion	n	Length	C.V.
Jackknife	100	0.0101010	0.0033839
	200	0.0050252	0.0013821
	400	0.0025063	0.0008130
	800	0.0012516	0.0004057
	1,609	**0.0006219**	**0.0001950**
Bootstrap	100	0.1100000	0.0322345
	200	0.0650000	0.0189570
	400	0.0575000	0.0159144
	800	0.0412500	0.0118012
	1,609	**0.0286047**	**0.0078527**

From Table 1, estimates of the proportion of Thais infected with COVID-19 Using the population size of 1609, the Jackknife method had a coefficient of variation (C.V.) of 0.01950% and the width of the confidence interval (Length) was 0.06219%. As for the bootstrap method, with 1000 repetitions, the coefficient of variation (C.V.) equal to 0.78527 percent and the confidence interval (Length) is 2.86047 percent.

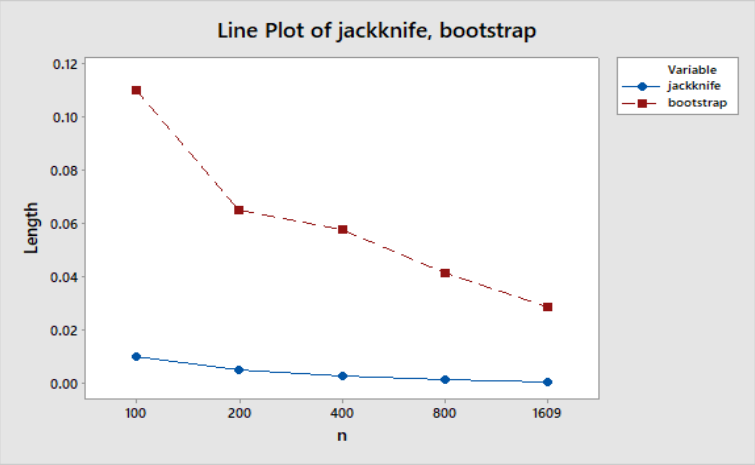

Figure 1 Length by Jackknifing Method and Bootstrapping Method

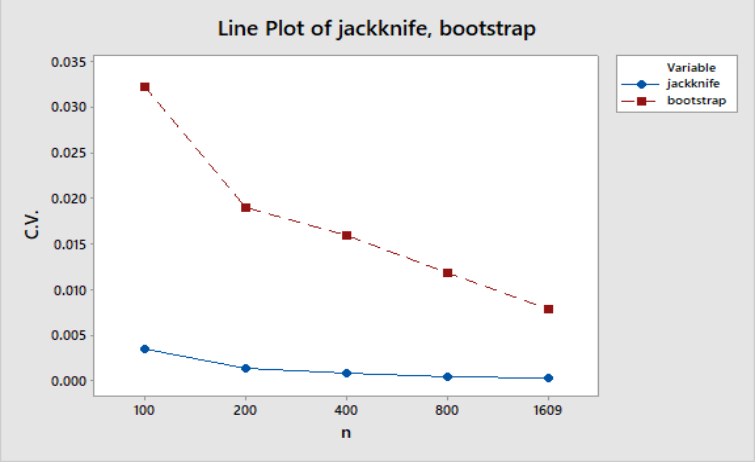

Figure 2 Coefficient of Variation or C.V. by Jackknifing Method and Bootstrapping Method

From Figure 1 and Figure 2, it was found that the width of the confidence interval. (Length) and coefficient of variation (C.V.) of the estimation of the proportion of Thai people infected with COVID-19. The Jackknife method was less than the width of the confidence interval and the coefficient of the variation of the proportional estimation. The bootstrap method of all sample sizes and as the sample size increases, the width of the confidence interval of both methods is reduced.

Estimation of proportion Classified by the independent variable, By the way, Jackknife and Bootstrap. Of all the data, when assigning the independent variables are gender, age, and region. Estimates the proportion of Thai people infected with COVID-19. By the Jackknife method and the bootstrap method. We will get the results of the analysis as shown in Table 2 – 4.

Table 2 Estimations of Proportion (COVID-19) for Thai in Thailand by sex

Estimate of Proportion	Sex	N	$\hat{p}$	Length	C.V.
Jackknife	Male	940	0.8765957	0.001065	0.00039979
	Female	669	0.9581465	0.001497	0.00031311
Bootstrap	Male	940	0.8764734	0.042606	0.01234290
	Female	669	0.9579611	0.029933	0.00806896

From Table 2, it was found that estimates of the proportion of Thais infected by COVID-19 by sex were 940 males and females 669 were found. Thai males are approximately 87.65957 percent, and Thai females are approximately 95.81465 percent and considering the width of the confidence interval. (Length) and coefficient of variation (C.V.) it was found that the Jackknife method had the width of the confidence interval and the coefficient of variation was less than the bootstrap method.

Table 3 Estimations of Proportion (COVID-19) for Thais in Thailand by age

Estimate of Proportion	Age	N	$\hat{p}$	Length	C.V.
Jackknife	< 18 years	36	0.9722222	0.0035714	0.00489796
	18–34 years	689	0.9245283	0.0014535	0.00041558
	35 - 60years	715	0.9090909	0.0014005	0.00044321
	> 60years	169	0.8461538	0.0059524	0.00254565
Bootstrap	< 18 years	36	0.9710833	0.0833333	.002945586
	18–34 years	689	0.9250174	0.0391872	0.01109709
	35 - 60years	715	0.9089958	0.0405594	0.01139269
	> 60years	169	0.8455740	0.1124261	0.03276909

From Table 3, estimates of the proportion of Thais infected with COVID-19 Classified by age the sample size under 18 years had a sample size of 36 people, 18-34 years of age, with a sample size of 689 people, 35-60 years of age, with a sample size of 715 people and age 60 years and over. There is a sample size of 169 people, the Jackknife method found that people under the age of 18 who were Tai had the highest estimates of the proportion of COVID-19 at 97.22222%, and the 18-34-year-old Thai population had an estimate of COVID-19 was followed by 92.45283%. When considering the width of the confidence interval, (Length) and value the coefficient of variation (C.V.) was found that the Jackknife method. Is the width of the span the confidence and coefficient of variation were lower than that of all age bootstraps.

Table 4 Estimations of Proportion (COVID-19) for Thai in Thailand by region

Estimate of Proportion	Region	N	$\hat{p}$	Length	C.V.
Jackknife	North	49	0.9387755	0.0208333	0.0053755
	Northeast	84	0.9642857	0.0111456	0.0023326
	Center	1177	0.9175871	0.0008503	0.0002549
	East	72	0.8888889	0.0140845	0.0050146
	West	27	0.7777778	0.0384615	0.0209502
	South	200	0.8650000	0.0050251	0.0019902
Bootstrap	North	49	0.9368776	0.1428571	0.0388790
	Northeast	84	0.9644881	0.0833333	0.0206817
	Center	1177	0.9175183	0.0305862	0.0087979
	East	72	0.8890139	0.1512777	0.0437350
	West	27	0.7808889	0.2962963	0.1024480
	South	200	0.8653650	0.0950000	0.0285338

From Table 4, it was found that the estimation of the proportion of Thais infected with COVID-19 was classified by region, with the North having a sample size of 49, the Northeast with

a sample size of 84, the Central region having a sample size of 1,177, the Eastern Region having the sample size. The sample size was 72, the western region had a sample size of 27, and the southern region had a sample size of 200. The Jack Knife method found that the Northeast of Thais had an estimate of the proportion of COVID-19. The most were 96.42857 percent and the North with Thai people has an estimated proportion of COVID-19, followed by 93.87755 percent. And when considering the width of the confidence interval (Length) and the coefficient of variation (C.V.), it was found that the Jackknife method had the width of the confidence interval and the coefficient of variation. It is less than the bootstrap method in all sectors.

4. SUMMARY AND DISCUSSION

This study provided a comparison of the estimate efficiency for the proportion of Thai people infected with the COVID-19 by both point estimation and interval estimation between the Jackknifing method and the Bootstrapping method. These methods were compared for their coefficients of variation (C.V.) and the widths of their confidence intervals (Length) The method with the lower value of the coefficient of variations and with lower confidence intervals can be concluded as the more efficient method. The results found that estimation of proportions by the Jackknifing method had a coefficient of the variation and width of the confidence interval lower than did the Bootstrap method for all sample sizes when the sample size increased, As a result, both the coefficient of variation and the width of the confidence were lower. Therefore, the Jackknifing method was more effective than Bootstrapping method.

As to estimates of the proportion of Thai people infected with COVID-19 by sex, age, and region, the results showed that the proportion of females to be higher than that of males in both the Jackknifing method and the Bootstrap method. Thai people under the age of 18 had the highest percentage of COVID-19 infections detected. The region with the highest percentage of COVID-19 detected was the Northeast region. These results were considered by the coefficient of the variation and the width of the confidence interval. They also indicated that the Jackknife method performed better than the Bootstrap method.

This research studied and compared with the unclassified data, the estimates of proportion

($\hat{p}$), the coefficients of variation and the widths of confidence intervals using the Jackknife and Bootstrap methods. In the future, researchers might conduct their research with both continuous and discrete data distributions, and they might study other estimations of parameters such as medians, varianc, etc.

CONFLICT OF INTERESTS

The author(s) declare that there is no conflict of interests.

REFERENCES

[1] K. Puttachaiyatad, Comparison of symmetrical and asymmetric distribution estimation with Jackknife method and bootstrap, Master's Thesis, Chulalongkorn University, 2009.

[2] J. Singpong, Estimation of the confidence interval of parameters of the extreme value distribution. when the data is censored, type 1. Master's Thesis, Thammasat University, 2017.

[3] J. Raksiharn, P. Thaninphong, C. Kongpan, L. Saenchan, Comparison of hypertensive prevalence estimates among residents of Doi Tao district. Chiang Mai province by the bootstrap method and Jackknife method, Master's Thesis, Chiang Mai University, 2014.

[4] Emergency operation center Department of Disease Control. 2020. Coronavirus Disease Situation 2019. Available: https://ddc.moph.go.th/viralpneumonia/situation.php, 18 August 2020.

[5] B. Efron, Bootstrap methods: another look at the Jackknife, in: S. Kotz, N.L. Johnson (Eds.), Breakthroughs in Statistics, Springer New York, New York, NY, 1992: pp. 569–593.

[6] B. Efron, Discussion: Jackknife, bootstrap and other resampling methods in regression analysis, Ann. Stat. 14 (1986), 1301-1304.