บทความวิจัย

ช่วงความเชื่อมั่นของความแปรปรวนด้วยวิธีโบเน็ตทีร่วมกับค่าเฉลี่ยเรขาคณิตสำหรับข้อมูลที่ไม่มีการแจกแจงปรกติ

อนุรักษ์ ทองขาว*
กลุ่มวิชาการเงินและคณิตศาสตร์ คณะการจัดการบริการ วิทยาลัยดุสิตธานี ศูนย์การศึกษาเมืองพัทยา

* ผู้นิพนธ์ประสานงาน โทรศัพท์ 08 3598 0543 อีเมล: anurak.to@dtc.ac.th DOI: 10.14416/j.kmutnb.2024.10.001
รับเมื่อ 24 เมษายน 2566 แก้ไขเมื่อ 11 กรกฎาคม 2566 ตอบรับเมื่อ 25 กรกฎาคม 2566 เผยแพร่ออนไลน์ 1 ตุลาคม 2567

บทคัดย่อ

การวิจัยนี้มีวัตถุประสงค์เพื่อพัฒนาช่วงความเชื่อมั่นของความแปรปรวนสำหรับข้อมูลประชากรกลุ่มเดียวที่ไม่มีการแจกแจงปรกติ และศึกษาประสิทธิภาพของช่วงความเชื่อมั่นของความแปรปรวนสำหรับข้อมูลประชากรเดียวที่ไม่มีการแจกแจงปรกติ ใช้เทคนิคมอนติคาร์โลทำซ้ำ 50,000 ครั้ง ในการจำลองข้อมูลขนาดต่าง ๆ ที่ไม่มีการแจกแจงปรกติ (ไค-สแควร์ เอกซ์โพเนนเชียล แกมมา และไวบูลล์) วิธีการประมาณค่าช่วงความเชื่อมั่นของความแปรปรวนสำหรับข้อมูลประชากรกลุ่มเดียวที่ไม่มีการแจกแจงปรกติที่ศึกษามี 3 วิธี คือ 1) วิธีของโบเน็ต 2) วิธีการปรับโบเน็ตทีร่วมกับค่ามัธยฐาน และ 3) วิธีการปรับโบเน็ตทีร่วมกับค่าเฉลี่ยเรขาคณิต โดยพิจารณาประสิทธิภาพจากค่าความน่าจะเป็นครอบคลุม และค่าความยาวเฉลี่ย ผลการวิจัยแสดงให้เห็นว่าเมื่อข้อมูลที่ไม่มีการแจกแจงปกติมีขนาดตัวอย่างที่เล็ก ช่วงความเชื่อมั่นของความแปรปรวนวิธีของโบเน็ตทีร่วมกับค่าเฉลี่ยเรขาคณิต มีประสิทธิภาพที่ดีกว่าวิธีอื่น ๆ

คำสำคัญ: ความแปรปรวน ช่วงความเชื่อมั่น ค่าเฉลี่ยเรขาคณิต

การอ้างอิงบทความ: อนุรักษ์ ทองขาว, “ช่วงความเชื่อมั่นของความแปรปรวนด้วยวิธีโบเน็ตทีร่วมกับค่าเฉลี่ยเรขาคณิตสำหรับข้อมูลที่ไม่มีการแจกแจงปรกติ,” *วารสารวิชาการพระจอมเกล้าพระนครเหนือ*, ปีที่ 35, ฉบับที่ 2, หน้า 1–10, เลขที่บทความ 252-136870, เม.ย.–มิ.ย. 2568.

วารสารวิชาการพระจอมเกล้าพระนครเหนือ ปีที่ 35 ฉบับที่ 2 เม.ย.–มิ.ย. 2568
The Journal of KMUTNB., Vol. 35, No. 2, Apr.–Jun. 2025

Research Article

The Confidence Interval of Variance by Adjusted Bonett-t with the Geometric Mean Method for Non-normal Distributions

Anurak Tongkaw*
Finance and Mathematics Cluster, Faculty of Hospitality Management, Dusit Thani College, Pattaya Campus, Chon Buri, Thailand

* Corresponding Author, Tel. 08 3598 0543, E–mail: Anurak.to@dtc.ac.th DOI: 10.14416/j.kmutnb.2024.10.001
Received 24 April 2023; Revised 11 July 2023; Accepted 25 July 2023; Published online: 1 October 2024

Abstract

The objectives of this research were 1) to develop a confidence interval for the one-population variance for non-normal distribution data and 2) to study the efficiency of the confidence interval developed for non-normal distribution data. The simulation was implemented by using 50,000 Monte Carlo random samples of a given sample size from various non-normal distributions (Chi-square, Exponential, Gamma, and Weibull). There were 3 methods for estimating the confidence interval of variance for the one-population variance of non-normal distribution data: 1) Bonet's method, 2) The adjusted Bonett-t with the Median method, and 3) The adjusting Bonett-t with the geometric mean method. The performance considerations of the confidence interval consisted of coverage probability and average length. The results showed that when the sample size was small and the data were not normally distributed, the adjusted Bonett-t with the geometric mean method performed better than the other methods.

Keywords: Confidence Interval, Variance, Geometric Mean

Please cite this article as: A. Tongkaw, "The confidence interval of variance by adjusted bonett-t with the geometric mean method for non-normal distributions," *The Journal of KMUTNB*, vol. 35, no. 2, pp. 1–10, ID. 252-136870, Apr.–Jun. 2025.

1. Introduction

Considering only the central tendency would not be enough to analyze data effectively. The distribution of data also should be under consideration. One popular value of statistics used to measure the distribution of data in many fields is variance (σ^2). If the variance has a high value, then the data is highly distributed, and if the variance has a low value, then the data is less distributed. [1] There are two types of parameters (population variance) estimation which are 1) point estimation and 2) interval estimation. The interval estimation can indicate the probability of the correctness of the parameter estimates. Therefore, interval estimation is more popular than point estimation. [2]

In some cases, the analyzed data is non-normal distributions for which the traditional estimation method of variance confidence intervals (χ^2 method) would not be appropriate because the traditional method was developed from the concept of normal distribution. In this case, a variance estimation method must be used for non- normal distribution. Bonett [3] proposed an interval estimate of the variance for non-normal distribution. This method provides a comprehensive probability that shows the effectiveness of the confidence interval estimation method. Many cases give good value. For example, a Beta distribution with sample sizes 25, 50, and 100 has coverage probability which is close to 0.95. However, there are still many cases that need to be developed. For instance, a Chi-square distribution with sample sizes 10, 25, 50, and 100 has coverage probability which is not close to 0.95. Niwitpong and Kerdwichai [4] developed a Bonett [3] variance estimation method. The median was used to assess the ferret. This is because there is a concept that when data is non-normal distribution the median is the central value that represents the data that is more effectively than the arithmetic mean.

The use of Student’s t distribution replacing standard normal distribution was included because the data is small sample size for non-normal distributions, the use of student T distribution for estimation of confidence intervals of one population variance was more efficient [5], [6].

The method developed by Niwitpong and Kerdvichai [4] is more efficient than the Bonett method [3] in many cases. However, the efficiency of that method can be improved because the coverage probability is not as good as it should be in many cases. The researcher is interested in using the geometric mean for the estimation of kurtosis modified by the Bonett method [3] because the geometric mean is suitable for highly dispersed data or the distance between the data value that is very different [7] and student T distribution according to the concept of Niwitpong and Kerdvichai.

The researcher expected that the new one-population variance confidence interval developed will perform better compared to the Bonett method [3] and Niwitpong and Kerdvichai method [4] in many cases for the non- normal distribution. Finally, the developed one-population variance confidence intervals are useful in data analysis in all disciplines. The objectives of this research were 1) to develop a confidence interval for the one-population variance for non-normal distribution data and 2) to study the efficiency of the confidence interval developed for non-normal distribution data.

2. Materials and Methods

In this research, the efficiency of one-population variance confidence interval estimation for non-normal distributions was studied by 3 methods, which consisted of 1) the Bonett method, 2) the adjusted Bonett-t with the Median method developed by Niwitpong and Kerdvichai [2], and the adjusted Bonett-t with the Geometric Mean method which is a new method developed. Let $x_1, x_2, x_3, \ldots x_n$ be a random sample of size n and x has non-normal distribution. All the variance confidence interval estimations in this study are as follows.

2.1 The confidence intervals of the one-population variance for non-normal distributions

2.1.1 The Bonett method

The Bonett method 100(1–∝)% confidence interval for the one-population variance (σ^2) is [3]

$$Exp\{ln(cs^2) \pm z_{1-\propto/2} se\} \quad (1)$$

where $z_{1-\propto/2}$ is the $(1-\propto/2)\,100^{th}$ percentile of the standard normal distribution,

$$se = c[\{\hat{\gamma}_4^*(n-3)/n\}/(n-1)]^{1/2} \quad (2)$$

$c = n/(n - z_{1-\propto/2})$ is a small-sample adjustment, $\hat{\gamma}_4^* = (n_0\tilde{\gamma}_4 + n\overline{\gamma}_4)/(n_0 + n)$, $\tilde{\gamma}_4$ is a prior estimate of γ obtained from a larger sample of size n_0, However, if $\tilde{\gamma}_4$ is not available then $\hat{\gamma}_4^* = \overline{\gamma}_4$ should be used. In the scope of this research, only such cases will be studied.

$$\overline{\gamma}_4 = \frac{n\sum_{i=1}^{n}(x_i - \overline{x}_{trim})^4}{\left(\sum_{i=1}^{n}(x_i - \overline{x})^2\right)^2} \quad (3)$$

where $\overline{x}_{trim}$ is a trimmed mean with trim-proportion equal to $1/\{2(n-4)^{1/2}\}$ so that m converges to μ as n increases without bound. This estimator of kurtosis tends to have less negative bias and smaller coefficient of variability than Pearson's estimator in symmetric and skewed leptokurtic distributions

2.1.2 The adjusted Bonett-t with the Median method

The adjusted Bonett-t with the median method $100(1-\propto)\%$ confidence interval for the one-population variance is [4]

$$Exp\{ln(c_t s^2) \pm t_{1-\propto/2,\vartheta}\, se_{Me}\} \quad (4)$$

where $t_{1-\propto/2}$ is the $(1-\propto/2)100^{th}$ percentile of the t-distribution with $\vartheta = n - 1$ degree of freedom and $c_t = n/(n - t_{1-\propto/2})$.

$$se_{Me} = c_t[\{\hat{\gamma}_{4Me}(n-3)/n\}/(n-1)]^{1/2} \quad (5)$$

$$\hat{\gamma}_{4Me} = \frac{n\sum_{i=1}^{n}(x_i - Me)^4}{\left(\sum_{i=1}^{n}(x_i - \overline{x})^2\right)^2} \quad (6)$$

where Me is a median which uses the kurtosis estimation developed follow concept by Niwitpong and Kerdvichai [4]

2.1.3 The adjusting Bonett-t with the geometric mean method

The adjusted Bonett-t with the Geometric Mean method $100(1-\propto)\%$ confidence interval for the one-population variance is

$$Exp\{ln(c_t s^2) \pm t_{1-\propto/2,\vartheta}\, se_{GM}\} \quad (7)$$

where $t_{1-\propto/2}$ is the $(1-\propto/2)100^{th}$ percentile of the t-distribution with $\vartheta = n - 1$ degree of freedom and $c_t = n/(n - t_{1-\alpha/2})$.

$$se_{GM} = c_t[\{\hat{\gamma}_{4GM}(n-3)/n\}/(n-1)]^{1/2} \quad (8)$$

$$\hat{\gamma}_{4GM} = \frac{n\sum_{i=1}^{n}(x_i - GM)^4}{\left(\sum_{i=1}^{n}(x_i - \bar{x})^2\right)^2} \quad (9)$$

where GM is a geometric mean ($GM = (\prod_{i=1}^{n} x_i)^{\frac{1}{n}}$) which uses the kurtosis estimation developed in this research.

2.2 Simulation Study

This section provides simulation studies for the coverage probabilities and the average lengths of confidence intervals of the one-population variance for non-normal distributions as proposed in section 2.1. The simulation samples were written in *r* programming using 50,000 Monte Carlo. The confidence level was 95%, samples of size (n) = 10, 25, 50, and 100 were generated from chi-square distribution, exponential distribution, Weibull distribution and gamma distribution. The distribution characteristics of the studies data are positive skewed.

The different confidence intervals of one-population variance for non-normal distribution, their estimated coverage probabilities and average length were considered. For each of the methods considered, a $(1-\alpha)100\%$ confidence interval denoted by (L, U) was obtained (based on $M = 50{,}000$ replicates). The estimated coverage probability (CP) and the average length (AL) are given by [3], [8].

$$CP = \frac{\sum_{j=1}^{M} Coverage_j}{M};\ j = 1, \ldots, 50{,}000 \quad (10)$$

where $Coverage_j$ has a value of 1 when the confidence intervals cover variances of one population and $Coverage_j$ has a value of 0 when the confidence intervals do not cover variances of one population.

$$AL = \frac{\sum_{j=1}^{M}(U_j - L_j)}{M};\ j = 1, \ldots, 50{,}000 \quad (11)$$

The (L_j, U_j) is the lower confidence interval and upper confidence interval. The performance considerations of the confidence interval (10) are determined by the coverage probabilities at closer to the confidence level ($1-\alpha$) and then compared with the average length (11). If any methods to compare the coverage probabilities value get close to the confidence level (0.95), then consider the width of the average length by considering the shortest value of the average length.

3. Results and Discussion

Estimates of coverage probabilities and average length widths of (1), (4), and (7) were obtained using 50,000 Monte Carlo random samples of a given sample size from various non-normal distributions.

The results are as follows. Table 1 and Figure 1 show that when the data was Chi-square distribution, the adjusted Bonett-t with the Geometric Mean method had a coverage probability value closer to 0.95 than other methods in case 10, 25, and 50. For the case of the sample size of 100, the adjusted Bonett-t with the Median method gave a coverage probability closer to 0.95 than other methods. Table 2 and Figure 2 show that when the data was Exponential distribution, the adjusted Bonett-t with the Geometric Mean method had a coverage probability value closer to 0.95 than other methods.

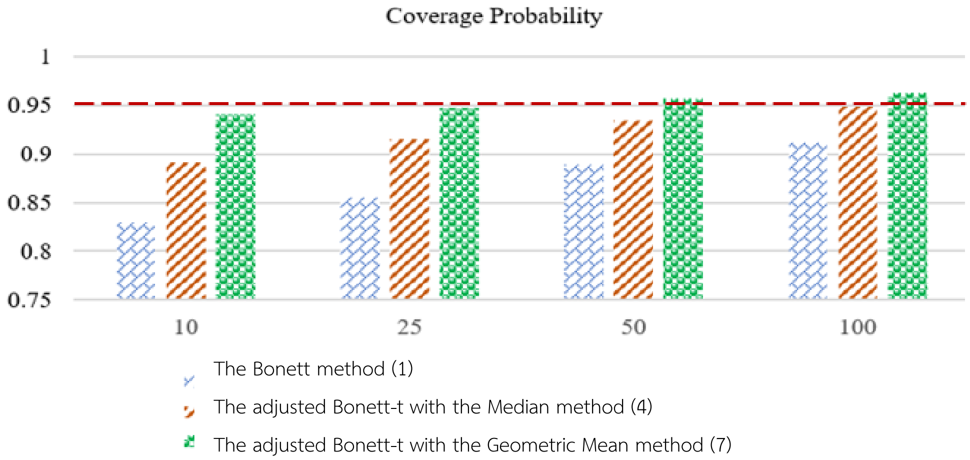

Figure 1 The coverage probability of 95% confidence intervals for Chi-square distribution.

Table 1 The estimated Coverage Probability (CP) and the Average Length (AL) of 95% confidence intervals for Chi-square distribution ($X \sim chisq(1)$, $Variance = 2$).

Confidence Intervals	*CP & AL*	10	25	50	100
The Bonett method (1)	*CP*	0.8435	0.8610	0.8901	0.9114
	AL	17.1795	7.0218	4.6126	3.1244
The adjusted Bonett-t with the Median method (4)	*CP*	0.895	0.9152	0.9349	**0.9484**
	AL	24.7178	9.0270	5.5626	3.6420
The adjusted Bonett-t with the Geometric Mean method (7)	*CP*	**0.9481**	**0.9484**	**0.9563**	0.9649
	AL	32.7509	10.2522	6.0679	3.9129

Note: Bold was closer to 0.95 than other methods.

For the case of the sample size of 100, the adjusted Bonett-t with the Median method gave a coverage probability closer to 0.95 than other methods. For the case of sample sizes of 50 and 100, the adjusted Bonett-t with the Median method gave a coverage probability closer to 0.95 than other methods.

Table 3 and Figure 3 show that when the data was Gamma distribution, the adjusted Bonett-t with the Geometric Mean method has a coverage probability value closer to 0.95 than other methods. For the case of the sample size of 100, the adjusted Bonett-t with the Median method gave a coverage probability closer to 0.95 than other methods. For the case of sample sizes of 50 and 100, the adjusted Bonett-t with the Median method gave a coverage probability closer to 0.95 than other methods. Table 4 and Figure 4 show that when the data was Weibull distribution, the adjusted Bonett-t with the Geometric Mean method had a coverage probability value closer to 0.95 than other methods.

For the case of the sample size of 100, the adjusted Bonett-t with the Median method gave a

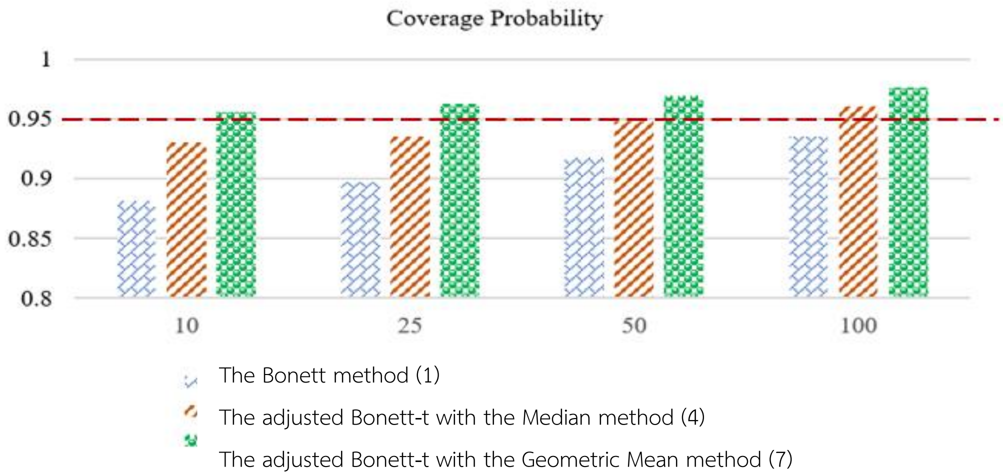

Figure 2 The coverage probability of 95% confidence intervals for Exponential distribution.

Table 2 The estimated Coverage Probability (CP) and the Average Length (AL) of 95% confidence intervals for Exponential distribution ($X \sim Exp(1)$, $Variance = 1$).

Confidence Intervals	*CP & AL*	10	25	50	100
The Bonett method (1)	*CP*	0.8905	0.8989	0.9193	0.9349
	AL	6.6312	2.7476	1.7890	1.2205
The adjusted Bonett-t with the Median method (4)	*CP*	0.9286	0.9356	**0.9480**	**0.9602**
	AL	8.8061	3.4466	2.1021	1.3970
The adjusted Bonett-t with the Geometric Mean method (7)	*CP*	**0.9627**	**0.9631**	0.9693	0.9742
	AL	11.6300	3.8459	2.3141	1.5216

Note: Bold was closer to 0.95 than other methods.

coverage probability closer to 0.95 than other methods. For the case of sample sizes of 50 and 100, the adjusted Bonett-t with the Median method gave a coverage probability closer to 0.95 than other methods.

When considering the coverage probability from all the research results, it can be seen that when the sample size is 10 and 25, known as the small sample size, the adjusted Bonett-t with the Geometric Mean method is the method with the probability value. It is closer to 0.95 than the other methods because when the sample size is small, the data is more spread out, and using the Geometric mean is a good approximation of the center of the data [7]. When considering the average length, as the sample size increases, the mean length gradually decreases, which is consistent with the research of Tongkaw [8] and Phuenaree and Sanorsap [9] which have the same characteristics.

This is because when the sample size is larger, the data is less distributed, making the estimation more accurate. The Bonett method gave the average length less than other methods or was too small,

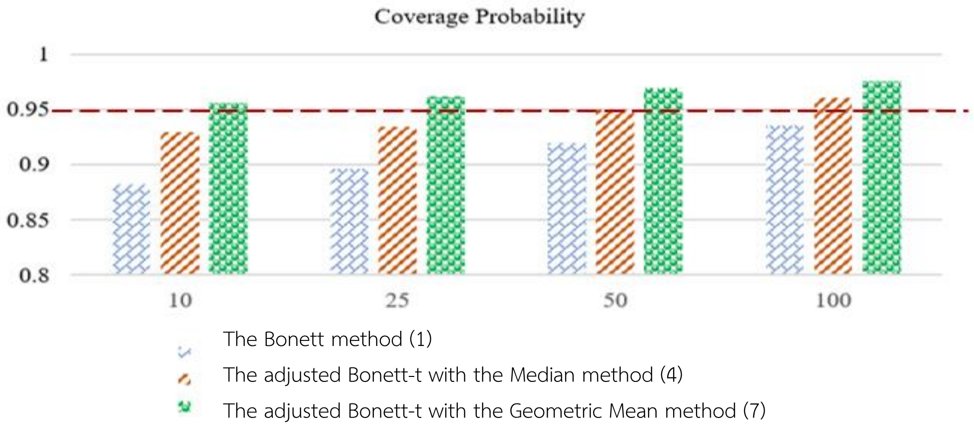

Figure 3 The coverage probability of 95% confidence intervals for Gamma Distribution.

Table 3 The estimated Coverage Probability (CP) and the Average Length (AL) of 95% confidence intervals for Gamma distribution ($X \sim Gamma(1,3)$, $Variance = 1/9$).

Confidence Intervals	*CP & AL*	10	25	50	100
The Bonett method (1)	*CP*	0.8969	0.8976	0.9182	0.9351
	AL	0.7359	0.3051	0.1978	0.1361
The adjusted Bonett-t with the Median method (4)	*CP*	0.9287	0.9345	**0.9486**	**0.9611**
	AL	0.9764	0.3782	0.2327	0.1558
The adjusted Bonett-t with the Geometric Mean method (7)	*CP*	**0.9631**	**0.9621**	0.9678	0.9764
	AL	1.2853	0.4274	0.2569	0.1692

Note: Bold was closer to 0.95 than other methods.

resulting in poor coverage probability. The adjusted Bonett-t with the Geometric Mean method and the adjusted Bonett-t with the Geometric Mean method had a greater mean length than the Bonett method, resulting in a coverage probability close to 0.95.

4. Conclusion

The confidence interval for one-population variance when the data has non-normal distributions developed in this research was the adjusted Bonett-t with the Geometric Mean method (7). The performance of the confidence intervals simulated data with the *R* program using the Manti Carlo method 50,000 times by studying the sample sizes 10, 25, 50, and 100 with non-normal distributions. A total of 16 cases were studied, of which 9 cases were the adjusted Bonett-t with the Geometric Mean method (7) and it was more effective than the other methods or 56.25% (9/16). The Geometric mean was used to improve confidence interval because Geometric mean was an appropriate measure of central tendency when the data is very dispersed [7], which

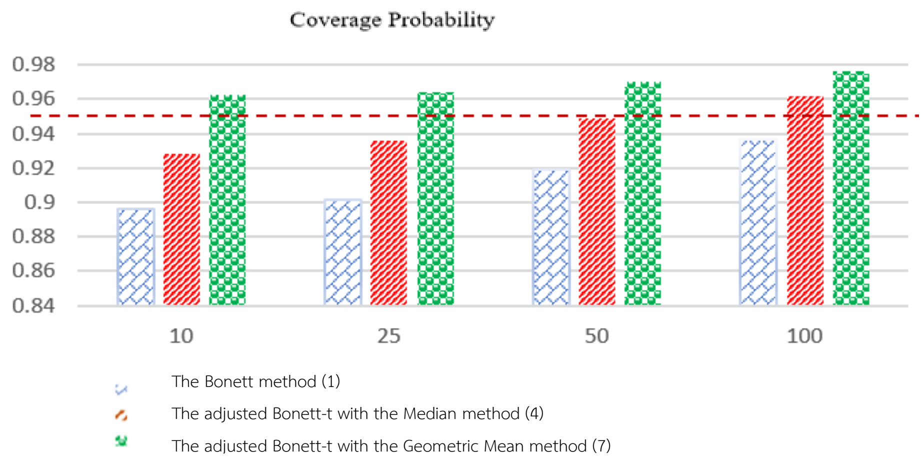

Figure 4 The coverage probability of 95% confidence intervals for Weibull Distribution.

Table 4 The estimated Coverage Probability (CP) and the Average Length (AL) of 95% confidence intervals for Weibull distribution (X~Weibull(1,1), $Variance = 1$).

Confidence Intervals	*CP & AL*	10	25	50	100
The Bonett method (1)	*CP*	0.8960	0.9016	0.9176	0.9369
	AL	6.7241	2.7577	1.7919	1.2241
The adjusted Bonett-t with the Median method (4)	*CP*	0.9283	0.9358	**0.9490**	**0.9615**
	AL	8.9180	3.4155	2.1070	1.4012
The adjusted Bonett-t with the Geometric Mean method (7)	*CP*	**0.9627**	**0.9628**	0.970	0.9765
	AL	11.6386	3.8530	2.3199	1.5222

Note: Bold was closer to 0.95 than other methods.

is consistent with the non-normal distribution in this study. However, in this study, the previous studies' scope of the kurtosis case with large sample size was unknown. In the future, further studies should be conducted in such cases or from other non-normal distribution data or use other estimations of kurtosis when the data is non-normal distributios.

Example for Application In an experiment on the development of lotus petal jam, 10 customers tasted and rated their satisfaction on taste, smell, texture, and other characteristics. The data obtained is the overall average of all aspects of each customer. The full score is 9 points. Reporting results need to report both the mean and the standard deviation. The data from the experiment is as follows.

When considering the Histogram of Figure 5, and statistical tests, it was found that the data was non-normal distribution (Mean = 7.41, S.D. = 1.25, Kurtosis = –0.12, and Skewness = 0.55). Therefore, the researcher uses the adjusted Bonett-t with the Geometric Mean method to estimate the standard deviation. The adjusted Bonett-t with the Geometric

9.75, 8.59, 6.66, 6.77, 7.23, 8.57, 6.50, 7.75, 6.78, 5.55

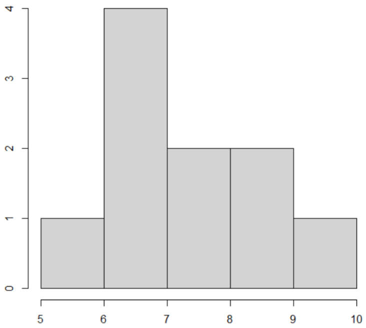

Figure 5 Histogram diagram of satisfaction data.

Mean method illustrates that 95% confidence interval for the one-population standard deviation ($\sqrt{\sigma^2} = \sigma$) of this case is (0.74, 2.61). In food or beverage development research in the hospitality industry, there are many instances where data looks like this example. For better and more reliable analytical performance to enhance the quality of product development of the service industry, the confidence interval estimation method for the variance obtained from this research should be applied.

References

[1] S. Pongwichai, *Statistical data analysis by computer: emphasizing for research*, 26th ed. Bangkok: Chulalongkorn University Press, 2015. (in Thai)

[2] K. Vanichbuncha, *Principles of Statistics*, 12th ed. Bangkok: Chulalongkorn University Press, 2010 (in Thai).

[3] D. G. Bonett, "Approximate confidence interval for standard deviation of nonnormal distributions," *Computational Statistics & Data Analysis*, vol. 50, no. 3, pp. 775–782, 2006.

[4] S. A. Niwitpong and P. Kirdwichai, "Adjusted Bonett confidence interval for Standard deviation of non-normal distributions," *Thailand Statistician*, vol. 6, no. 1, pp. 1–16, 2008.

[5] W. Pan, "Approximate confidence intervals for one proportion and difference of two proportions," *Computational statistics & data analysis*, vol. 40, no. 1, pp. 143–157, 2002.

[6] U. Olsson, "Confidence intervals for the mean of a log-normal distribution," *Journal of Statistics Education*, vol. 13, no. 1, pp.1–9, 2005.

[7] A. Sakworawich, *Descriptive Statistics, Exploratory Data Analysis, and Statistical Graphic with Basic R Programming*, 2th ed. Bangkok: Chulalongkorn University Press, 2017 (in Thai).

[8] A. Tongkaw, "Confidence intervals for a ratio of two population medians by price and bonett bootstrap-t method," *The Journal of KMUTNB*, vol. 32, no. 2, pp. 457–468, 2022(in Thai).

[9] B. Phuenaree, & S. Sanorsap, "An Interval Estimation of Pearson's Correlation Coefficient by Bootstrap Methods," *Asian Journal of Applied Sciences*, vol. 5, no. 3, pp. 623–627, 2017.