

On Performance of Confidence Interval Estimate of Mean for Skewed Populations: Evidence from Examples and Simulations

Khairul Islam^{1*} and Tanweer J Shapla²

^{1,2}Department of Mathematics and Statistics

Eastern Michigan University, Ypsilanti, MI 48197, USA

*E-mail of the corresponding author: kislam@emich.edu

Abstract

The performances of confidence interval (CI) estimates of mean for skewed distributions are compared for three traditional methods and two newly proposed methods using coverage probability and confidence length for varying levels of skewness via simulations. Two real-life examples are incorporated to justify the applicability of the two newly proposed methods (trimmed t and modified trimmed t CIs), compared to the traditional methods (Student's t mad t and median t CIs). From the results of examples and simulation study, it appears that with skewed distribution, the proposed trimmed t and modified trimmed t CIs are as good as mad t or median t CIs in coverage probability consideration. With lower % trimmed, trimmed and modified trimmed t CIs are identical or close to the Student's t CI, and with increased % trimmed, they are identical or close to the median t CI.

Keywords: Student's t , Mad t , Median t , Modified trimmed t , Coverage probability, Length of confidence interval.

1. Introduction

Let X_1, X_2, \dots, X_n be a random sample from any skewed distribution with mean μ and standard deviation σ . Given the sample, we wish to find the confidence interval (CI) for μ when the population standard deviation σ is unknown. The sample mean \bar{X} of a random sample for any population with mean μ and standard deviation σ is approximately distributed as normal with a mean μ and standard deviation σ/\sqrt{n} , provided n is large. Therefore, when σ is known, the statistic $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ follows a standard normal distribution. As such, a $100(1-\alpha)\%$ CI for μ is given by

$$\left[\bar{X} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right]$$

where, $z_{\alpha/2}$ is the upper $(\alpha/2)$ th percentile of the standard normal distribution.

In real life, however, it is unlikely that σ is known. Then, an estimate of σ given by the sample standard deviation $s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2}$ is used to compute various t confidence intervals. Among various modifications, student's t (Student, 1908) CI is the most efficient and useful at normal models. Johnson (1978), proposed a modification of the Student's t CI for skewed distributions. Since Johnson (1978), Kleijnen et al. (1986), Meeden (1999), Willink (2005), Kibria (2006), Shi and Kibria (2007) are a few to mention who proposed several modifications. In this article, we proposed two methods of CIs for the mean of skewed populations, namely, trimmed t and modified trimmed t CIs.

The organization of the remaining paper is as follows. Student's t and various modifications appear in section 2. The proposed new methods of CIs are given in section 3. Two real life examples have been incorporated in section 4 to demonstrate applications of the new methods in relation to the other methods. A simulation study has been carried out in section 5 in order to compare performance of underlying CIs along with the proposed methods in terms of coverage probability and confidence length. Finally, a concluding remark is provided in section 6.

2. Various t Confidence Intervals (CIs)

In this section, we considered various versions of t CIs that are in practice when the population standard deviation σ is unknown.

2.1 Student's t CI

When the sample size n is small, the $100(1 - \alpha)\%$ CI for μ is due to Student (1908) and is given by

$$\left[\bar{X} - t_{\alpha/2, n-1} \frac{s}{\sqrt{n}}, \bar{X} + t_{\alpha/2, n-1} \frac{s}{\sqrt{n}} \right] \quad (1)$$

where $t_{\alpha/2, n-1}$ is the upper $\alpha/2$ percentage point of the Student's t distribution with $(n - 1)$ degrees of freedom. This CI is the most popular CI in literature and is omnipresent in statistical practice for making inference due to the efficiency of the method at normal models.

However, it is well known that when the population the sample comes from is skewed, Student's t CI has poor coverage property. In such a case Johnson's t (1978) along with several versions of modifications are available for practice

2.2 Johnson's t CI

When the sample size n is small and population distribution is non-normal or skewed, the Student's t CI has poor coverage probability. Johnson (1978) proposed the following CI for mean μ for a skewed distribution:

$$[\bar{X} + (\hat{\mu}_3/6s^2n)] \mp t_{\alpha/2, n-1} \frac{s}{\sqrt{n}} \quad (2)$$

where $\hat{\mu}_3 = \frac{n}{(n-1)(n-2)} \sum_{i=1}^n (X_i - \bar{X})^3$ is the unbiased estimator of the third central moment μ_3 .

It appears in literature (see for example, Kibria, 2006) that the width of Student's t and Johnson's t are same.

2.3 Median t CI

It is well known that \bar{X} is preferable to other estimators of centers for a distribution that is symmetric or relatively homogeneous. When the distribution is skewed or non-normal, the sample median describes the center of the distribution better than that of the mean. Therefore, for a skewed distribution, it is reasonable to define the standard deviation in terms of the median than the mean (Kibria, 2006). They proposed a new CI for μ by

$$\left[\bar{X} - t_{\alpha/2, n-1} \frac{\tilde{s}_1}{\sqrt{n}}, \bar{X} + t_{\alpha/2, n-1} \frac{\tilde{s}_1}{\sqrt{n}} \right] \quad (3)$$

where $\tilde{s}_1 = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \tilde{x})^2}$ and \tilde{x} is the sample median.

This CI they refer to as a median t CI.

2.4 Mad t CI

Kibria (2006) proposed another t CI which has been referred to as mad t CI.

A $100(1 - \alpha)\%$ mad t CI for μ is given by

$$\left[\bar{X} - t_{\alpha/2, n-1} \frac{\tilde{s}_2}{\sqrt{n}}, \bar{X} + t_{\alpha/2, n-1} \frac{\tilde{s}_2}{\sqrt{n}} \right] \quad (4)$$

where

$\tilde{s}_2 = \frac{1}{n} \sum_{i=1}^n |X_i - \tilde{x}|$ is the sample mean absolute deviation (MAD).

The Median t and Mad t CIs are ad-hoc types of CIs of μ for skewed distribution, which have also been considered by Shi and Kibria (2007). Merits of these CIs in comparison with Johnson's t interval have been shown by simulation study and examples.

3. New proposed t CIs

In between mean and median, the trimmed mean is a more robust measure for describing the center than the mean and more efficient than the median. We thought that for a skewed distribution with a longer left or right tail, it is reasonable to define the standard deviation in terms of the trimmed mean than mean or median. Therefore, we propose a modification of the Students' t CI given by

$$\left[\bar{X} - t_{\alpha/2, n-1} \frac{s_1}{\sqrt{n}}, \bar{X} + t_{\alpha/2, n-1} \frac{s_1}{\sqrt{n}} \right] \quad (5)$$

where $s_1 = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_{(p)})^2}$ and $\bar{X}_{(p)}$ is the trimmed mean with $p\%$ data values in both tails trimmed.

Another $100(1 - \alpha)\%$ t CI for μ is given by

$$\left[\bar{X} - t_{\alpha/2, n-1} \frac{s_2}{\sqrt{n}}, \bar{X} + t_{\alpha/2, n-1} \frac{s_2}{\sqrt{n}} \right] \quad (6)$$

where

$$s_2 = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \hat{\mu})^2}$$

$$\hat{\mu} = \begin{cases} \bar{X} & \text{if } X_{[np]} < \bar{X} < X_{[n(1-p)]} \\ \bar{X}_{(p)} & \text{other wise} \end{cases}$$

The two CIs in (5) and (6), we refer to as trimmed t and modified trimmed t confidence intervals. These are ad-hoc types of CIs of μ for skewed distribution, similar to Kibria (2006). We assess their performance by examples and simulations.

4. Examples

In this section, we provide two real-life examples in order to illustrate and compare performance of the two proposed trimmed and modified trimmed t CIs in relation to the existing popular alternatives, Students' t , med t and mad t CIs, when the samples are assumed to come from skewed distributions.

Example 4.1

Individuals with phenylketonuria (PKU) disorder are unable to metabolize the protein phenylalanine. In medical research, it has been suggested that an elevated level of serum phenylalanine increases a child likelihood of mental deficiency. The normalized mental age (nMA) score (in months) of a sample of 18 children is considered below from a population of children with high exposure of PKU disorder in order to assess the extent of their mental deficiency (see Wrona, R.M., 1979).

28, 35, 37, 37, 43.5, 44, 45.5, 46, 48, 48.3, 48.7, 51, 52, 53, 53, 54, 54, 55

We are interested to determine the 95% CI of mean normalized mental age score of children with high form of phenylketonuria.

From the histogram and boxplot in Figure 1 of the sample nMA score, it appears the population the sample comes from is a negatively skewed population. The sample mean and the sample skewness of this data are 46.3 and -0.98, respectively. From the t test ($t = 0.1536$, $df = 17$, $p\text{-value} = 0.8797$) and Wilcoxon signed rank test ($w = 83.5$, $p\text{-value} = 0.7581$), it is evident that the population data has the mean $\mu = 46$ months.

The 95% CIs together with the length of the corresponding CIs for this example are reported in Table 1.

Figure 1: Histogram and boxplot of the normalized mental age (nMA) score (in months) for the sample of children with higher form of phenylketonuria.

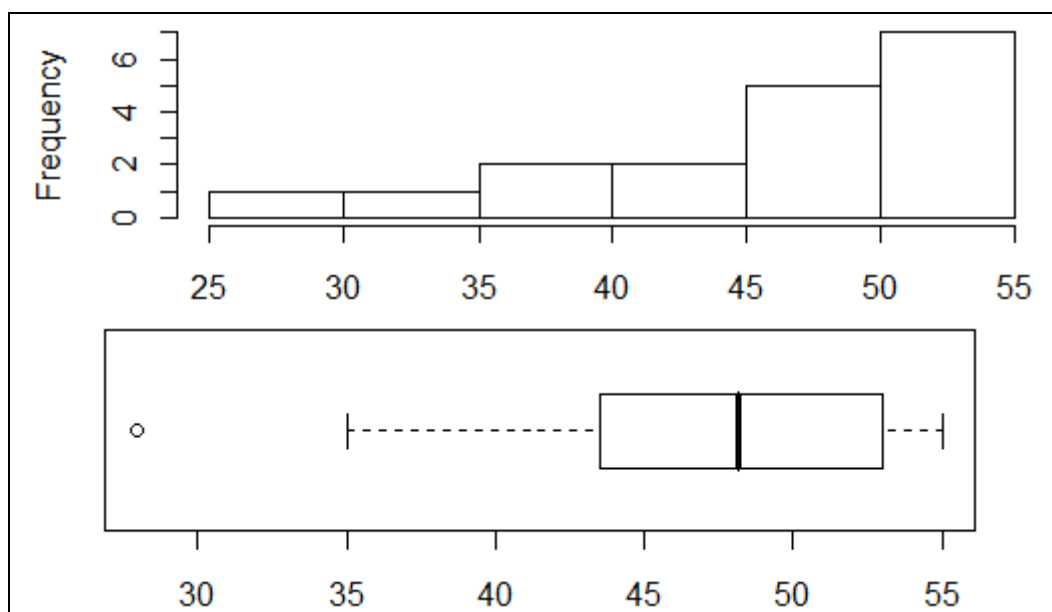


Table 1: 95% CIs with corresponding lengths for Example 1

% trimmed	Methods	CI	Length
	Student's t	(42.46,50.09)	7.63
	Median t	(42.34,50.21)	7.87
	Mad t	(43.28,49.27)	5.99
5% trimmed	Trimmed t	(42.46,50.09)	7.63
	Modified trimmed t	(42.46,50.09)	7.63
10% trimmed	Trimmed t	(42.45,50.11)	7.66
	Modified trimmed t	(42.46,50.09)	7.63
20% trimmed	Trimmed t	(42.41,50.14)	7.73
	Modified trimmed t	(42.46,50.09)	7.63
25% trimmed	Trimmed t	(42.36,50.19)	7.83
	Modified trimmed t	(42.46,50.09)	7.63

As we see from the 95% CIs reported in Table 1, all methods have captured the hypothesized mean $\mu = 46$. Lengthwise, Mad t CI has the shortest length (5.99). The student's t and Modified trimmed t have the second shortest length (7.63), following trimmed t and the median t , in order, respectively. By increasing the % trimmed, trimmed t CIs approach to med t CI. Modified trimmed t CI, retains the efficiency of Student's t and robustness of median t CIs.

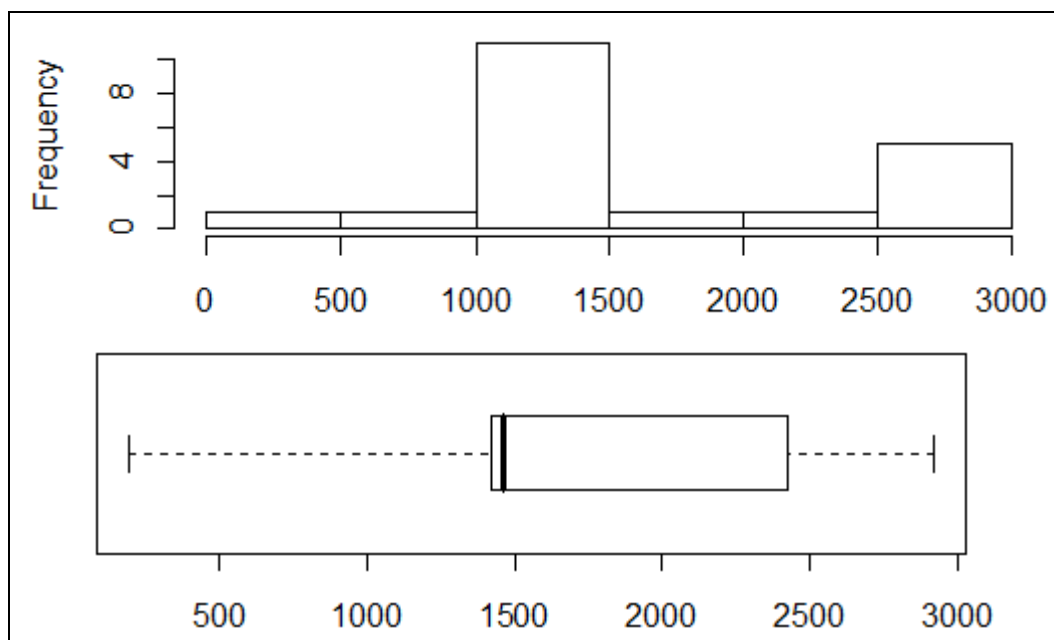
Example 4.2

A sample of size 20 is considered from the population of the number of days past presidents of the United States served in the office for the 43 Presidents as of 4 February 2004 (see Hayden, 2005). So the population has 43 data points with mean $\mu = 1824$ days and skewness=0.55. Therefore, the population is positively skewed. The sample data points are as follows:

2921, 1036, 2921, 1460, 1460, 2810, 1460, 881, 1418, 2810,
 1460, 1460, 199, 1503, 1110, 1418, 1461, 2921, 1460, 2039

From the sample, the point estimates of mean and skewness are 1710 days and 0.42, respectively. The histogram and boxplot in Figure 2 suggest that the sample comes from the population that is positively skewed.

Figure 2: Histogram and boxplot of the number of days US president served in the office in the sample.



The 95% CIs together with the length of the corresponding CIs for this example are reported in Table 2.

On the basis of 95% CI estimates reported in Table 2, all methods have captured the population mean $\mu = 1824$ days. Lengthwise, Mad t has the shortest length (163). Again, the student's t and Modified trimmed t have the second shortest length (724), following trimmed t and the median t , in order, respectively. With lower % trimmed (5% and 10%), the trimmed t and modified trimmed t CIs are identical to the Student's t CI. By increasing the % trimming, trimmed t CIs approach the med t CI. Overall, the modified trimmed t CI retains the efficiency of Student's t and robustness of median t CIs.

Table 2: 95% CIs with corresponding lengths for Example 2

% trimmed	Methods	CI	Length
	Student's t	(1348, 2072)	724
	Median t	(1329, 2092)	763
	Mad t	(1422, 1999)	577
5% trimmed	Trimmed t	(1348, 2072)	724
	Modified trimmed t	(1348, 2072)	724
10% trimmed	Trimmed t	(1348, 2072)	724
	Modified trimmed t	(1348, 2072)	724
20% trimmed	Trimmed t	(1346, 2075)	729
	Modified trimmed t	(1348, 2072)	724
25% trimmed	Trimmed t	(1337, 2084)	747
	Modified trimmed t	(1348, 2072)	724

5. Simulation and Result Discussion

In this section, we carry out a simulation study to compare the finite sample performance of the various CIs described in this article. All simulations are performed by using the statistical software R. The sample X is simulated from $G(\theta_1, \theta_2)$ population, where θ_1 is the shape parameter and θ_2 is the scale parameter. Note that the skewness of $G(\theta_1, \theta_2)$ distribution is $\gamma_1 = 2/\sqrt{\theta_1}$. In simulations, we choose different values of the parameter θ_1 to allow varying levels of skewness of the simulated samples, and the population mean is fixed at 1. In all simulations, the Monte Carlo size is 5,000, chosen arbitrarily.

The coverage probability of various CIs is estimated from the proportion of CIs containing the true mean 1 over all MC simulations. While considered trimmed and modified trimmed t CIs, 5%, 10%, 20%, 30% and 45% data values are trimmed from both tails.

All computations of this article are performed using software R (2016).

Table 3 below provides the characteristics of various population models used in the simulation study.

Table 3 Values of α and γ_1 used in simulations of X

Models	θ_1	θ_2	γ_1	mean
M1	16	0.0625	0.5	1
M2	4	0.25	1	1
M3	1	1	2	1
M4	0.25	4	4	1

The performances of the simulations in terms of coverage probability are reported in Tables 4.1-4.6, while confidence length of 95% CIs are reported in Tables 5.1-5.6. The summary, minimum (min) and maximum (max) coverage probability for all 95% CIs are reported in for Table 4.6, and the confidence length in Table 5.6.

Table 4.1: Coverage probability of 95% CIs when skewness=0.50

n	t	Med	mad	Trimmed t CI (%)					Modified Trimmed t CI (%)				
				5%	10%	20%	30%	45%	5%	10%	20%	30%	45%
5	0.95	0.95	0.89	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95
10	0.95	0.95	0.89	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95
15	0.95	0.95	0.89	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95
20	0.96	0.96	0.89	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96
25	0.95	0.96	0.89	0.95	0.95	0.96	0.96	0.96	0.95	0.95	0.95	0.95	0.96
30	0.96	0.96	0.90	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96
35	0.95	0.95	0.88	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95
40	0.95	0.95	0.88	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95
45	0.95	0.95	0.88	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95
50	0.96	0.96	0.88	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96

Table 4.2: Coverage probability of 95% CIs when skewness=1

n	t	med	mad	Trimmed t CI (%)					Modified Trimmed t CI (%)				
				5%	10%	20%	30%	45%	5%	10%	20%	30%	45%
5	0.93	0.94	0.88	0.93	0.93	0.94	0.94	0.94	0.93	0.93	0.93	0.93	0.94
10	0.93	0.94	0.87	0.93	0.93	0.93	0.93	0.94	0.93	0.93	0.93	0.93	0.93
15	0.93	0.94	0.87	0.93	0.93	0.94	0.94	0.94	0.93	0.93	0.93	0.93	0.94
20	0.94	0.94	0.87	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94
25	0.95	0.95	0.88	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95
30	0.94	0.94	0.87	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94
35	0.94	0.95	0.87	0.94	0.94	0.95	0.95	0.95	0.94	0.94	0.94	0.94	0.95
40	0.95	0.95	0.88	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95
45	0.94	0.95	0.88	0.94	0.95	0.95	0.95	0.95	0.94	0.94	0.94	0.94	0.95
50	0.94	0.94	0.87	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94

Table 4.3: Coverage probability of 95% CIs when skewness=2

n	t	med	mad	Trimmed t CI (%)					Modified Trimmed t CI (%)				
				5%	10%	20%	30%	45%	5%	10%	20%	30%	45%
5	0.88	0.89	0.83	0.88	0.88	0.88	0.88	0.89	0.88	0.88	0.88	0.88	0.89
10	0.90	0.91	0.85	0.90	0.90	0.91	0.91	0.91	0.90	0.90	0.90	0.90	0.91
15	0.91	0.92	0.84	0.91	0.91	0.91	0.91	0.92	0.91	0.91	0.91	0.91	0.92
20	0.92	0.93	0.85	0.92	0.92	0.92	0.93	0.93	0.92	0.92	0.92	0.92	0.93
25	0.92	0.93	0.85	0.92	0.93	0.93	0.93	0.93	0.92	0.92	0.92	0.92	0.93
30	0.93	0.94	0.85	0.93	0.93	0.93	0.93	0.94	0.93	0.93	0.93	0.93	0.93
35	0.93	0.94	0.84	0.93	0.93	0.93	0.93	0.94	0.93	0.93	0.93	0.93	0.94
40	0.94	0.94	0.85	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94
45	0.93	0.94	0.85	0.93	0.93	0.93	0.94	0.94	0.93	0.93	0.93	0.93	0.94
50	0.94	0.95	0.85	0.94	0.94	0.94	0.95	0.95	0.94	0.94	0.94	0.94	0.95

Table 4.4: Coverage probability of 95% CIs when skewness=4

n	t	med	mad	Trimmed t CI (%)					Modified Trimmed t CI (%)				
				5%	10%	20%	30%	45%	5%	10%	20%	30%	45%
5	0.73	0.75	0.68	0.73	0.73	0.74	0.74	0.75	0.73	0.73	0.73	0.74	0.75
10	0.80	0.82	0.73	0.80	0.81	0.81	0.81	0.82	0.80	0.80	0.80	0.80	0.82
15	0.83	0.85	0.75	0.83	0.83	0.84	0.84	0.85	0.83	0.83	0.83	0.83	0.85
20	0.85	0.87	0.76	0.85	0.86	0.86	0.87	0.87	0.85	0.85	0.85	0.86	0.87
25	0.86	0.88	0.76	0.86	0.86	0.87	0.87	0.88	0.86	0.86	0.86	0.86	0.88
30	0.87	0.89	0.76	0.87	0.88	0.89	0.89	0.89	0.87	0.87	0.87	0.88	0.89
35	0.89	0.90	0.77	0.89	0.89	0.90	0.90	0.90	0.89	0.89	0.89	0.89	0.90
40	0.89	0.91	0.77	0.89	0.89	0.90	0.90	0.91	0.89	0.89	0.89	0.89	0.91
45	0.90	0.92	0.77	0.90	0.90	0.91	0.91	0.92	0.90	0.90	0.90	0.90	0.92
50	0.90	0.92	0.77	0.90	0.90	0.91	0.91	0.92	0.90	0.90	0.90	0.90	0.92

Table 4.5: Coverage probability of 95% CIs when skewness=8

n	t	med	mad	Trimmed t CI (%)					Modified Trimmed t CI (%)				
				5%	10%	20%	30%	45%	5%	10%	20%	30%	45%
5	0.48	0.50	0.45	0.48	0.48	0.49	0.49	0.50	0.48	0.48	0.48	0.49	0.50
10	0.60	0.61	0.53	0.60	0.60	0.61	0.61	0.61	0.60	0.60	0.60	0.60	0.61
15	0.65	0.66	0.56	0.65	0.65	0.66	0.66	0.66	0.65	0.65	0.66	0.66	0.66
20	0.69	0.70	0.58	0.69	0.70	0.70	0.70	0.70	0.69	0.69	0.70	0.70	0.70
25	0.73	0.74	0.59	0.73	0.73	0.74	0.74	0.74	0.73	0.73	0.74	0.74	0.74
30	0.73	0.74	0.57	0.74	0.74	0.74	0.74	0.74	0.73	0.74	0.74	0.74	0.74
35	0.77	0.78	0.59	0.77	0.77	0.78	0.78	0.78	0.77	0.77	0.77	0.78	0.78
40	0.79	0.80	0.59	0.79	0.79	0.80	0.80	0.80	0.79	0.79	0.79	0.80	0.80
45	0.80	0.80	0.58	0.80	0.80	0.80	0.80	0.80	0.80	0.80	0.80	0.80	0.80
50	0.79	0.80	0.57	0.80	0.80	0.80	0.80	0.80	0.79	0.79	0.80	0.80	0.80

Table 4.6: Minimum (min) and maximum (max) coverage probability of various 95% CIs for varying values of skewness and % trimming

Skewness=0.5													
				Trimmed t CI (%)					Modified Trimmed t CI (%)				
n	t	med	mad	5%	10%	20%	30%	45%	5%	10%	20%	30%	45%
min	0.95	0.95	0.88	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95
max	0.96	0.96	0.90	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96
Skewness=1													
min	0.93	0.94	0.87	0.93	0.93	0.93	0.93	0.94	0.93	0.93	0.93	0.93	0.93
max	0.95	0.95	0.88	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95
Skewness=2													
min	0.88	0.89	0.83	0.88	0.88	0.88	0.88	0.89	0.88	0.88	0.88	0.88	0.89
max	0.94	0.95	0.85	0.94	0.94	0.94	0.95	0.95	0.94	0.94	0.94	0.94	0.95
Skewness=4													
min	0.73	0.75	0.68	0.73	0.73	0.74	0.74	0.75	0.73	0.73	0.73	0.74	0.75
max	0.90	0.92	0.77	0.90	0.90	0.91	0.91	0.92	0.90	0.90	0.90	0.90	0.92
Skewness=4													
min	0.48	0.50	0.45	0.48	0.48	0.49	0.49	0.50	0.48	0.48	0.48	0.49	0.50
max	0.80	0.80	0.59	0.80	0.80	0.80	0.80	0.80	0.80	0.80	0.80	0.80	0.80

Table 5.1: Confidence length of 95% CIs when skewness=0.50

				Trimmed t CI (%)					Modified Trimmed t CI (%)				
n	t	med	mad	5%	10%	20%	30%	45%	5%	10%	20%	30%	45%
5	0.58	0.61	0.44	0.58	0.58	0.59	0.59	0.61	0.58	0.58	0.58	0.58	0.61
10	0.35	0.36	0.27	0.35	0.35	0.35	0.35	0.36	0.35	0.35	0.35	0.35	0.35
15	0.27	0.28	0.21	0.27	0.27	0.27	0.27	0.28	0.27	0.27	0.27	0.27	0.28
20	0.23	0.23	0.18	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23
25	0.20	0.21	0.16	0.20	0.20	0.21	0.21	0.21	0.20	0.20	0.20	0.20	0.21
30	0.18	0.19	0.15	0.18	0.19	0.19	0.19	0.19	0.18	0.18	0.18	0.18	0.19
35	0.17	0.17	0.13	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17
40	0.16	0.16	0.13	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16
45	0.15	0.15	0.12	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.15
50	0.14	0.14	0.11	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14

Table 5.2: Confidence length of 95% CIs when skewness=1

				Trimmed t CI (%)					Modified Trimmed t CI (%)				
n	t	med	mad	5%	10%	20%	30%	45%	5%	10%	20%	30%	45%
5	1.14	1.21	0.87	1.14	1.14	1.17	1.17	1.21	1.14	1.14	1.14	1.15	1.21
10	0.68	0.70	0.53	0.68	0.69	0.69	0.69	0.70	0.68	0.68	0.68	0.68	0.70
15	0.54	0.55	0.42	0.54	0.54	0.54	0.54	0.55	0.54	0.54	0.54	0.54	0.55
20	0.46	0.47	0.36	0.46	0.46	0.46	0.46	0.47	0.46	0.46	0.46	0.46	0.47
25	0.41	0.42	0.32	0.41	0.41	0.41	0.41	0.42	0.41	0.41	0.41	0.41	0.41
30	0.37	0.38	0.29	0.37	0.37	0.37	0.37	0.37	0.37	0.37	0.37	0.37	0.37
35	0.34	0.35	0.26	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34
40	0.32	0.32	0.25	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32
45	0.30	0.30	0.23	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30
50	0.28	0.29	0.22	0.28	0.28	0.28	0.28	0.29	0.28	0.28	0.28	0.28	0.29

Table 5.3: Confidence length of 95% CIs when skewness=2

n	t	med	mad	Trimmed t CI (%)					Modified Trimmed t CI (%)				
				5%	10%	20%	30%	45%	5%	10%	20%	30%	45%
5	2.14	2.31	1.62	2.14	2.14	2.21	2.21	2.31	2.14	2.14	2.14	2.18	2.30
10	1.33	1.40	1.00	1.33	1.35	1.37	1.37	1.40	1.33	1.33	1.33	1.34	1.40
15	1.04	1.10	0.79	1.04	1.05	1.07	1.07	1.10	1.04	1.04	1.05	1.05	1.10
20	0.90	0.95	0.67	0.91	0.91	0.93	0.93	0.95	0.90	0.90	0.90	0.90	0.95
25	0.80	0.84	0.59	0.80	0.81	0.82	0.82	0.84	0.80	0.80	0.80	0.80	0.84
30	0.73	0.76	0.54	0.73	0.74	0.75	0.75	0.76	0.73	0.73	0.73	0.73	0.76
35	0.67	0.70	0.50	0.67	0.68	0.69	0.69	0.70	0.67	0.67	0.67	0.67	0.70
40	0.63	0.66	0.47	0.63	0.63	0.64	0.65	0.66	0.63	0.63	0.63	0.63	0.66
45	0.59	0.61	0.44	0.59	0.59	0.60	0.60	0.61	0.59	0.59	0.59	0.59	0.61
50	0.56	0.58	0.41	0.56	0.57	0.57	0.58	0.58	0.56	0.56	0.56	0.56	0.58

Table 5.4: Confidence length of 95% CIs when skewness=4

n	t	med	mad	Trimmed t CI (%)					Modified Trimmed t CI (%)				
				5%	10%	20%	30%	45%	5%	10%	20%	30%	45%
5	3.59	4.03	2.67	3.59	3.59	3.83	3.83	4.03	3.59	3.59	3.59	3.77	4.03
10	2.32	2.55	1.63	2.32	2.38	2.46	2.46	2.55	2.32	2.32	2.35	2.40	2.55
15	1.90	2.08	1.29	1.90	1.93	2.01	2.01	2.08	1.90	1.90	1.93	1.96	2.08
20	1.65	1.80	1.09	1.67	1.70	1.74	1.76	1.80	1.65	1.65	1.67	1.71	1.80
25	1.51	1.65	0.99	1.52	1.54	1.60	1.61	1.65	1.51	1.51	1.53	1.56	1.65
30	1.37	1.49	0.89	1.38	1.41	1.45	1.46	1.49	1.37	1.37	1.38	1.41	1.49
35	1.28	1.40	0.83	1.29	1.31	1.35	1.36	1.40	1.28	1.28	1.29	1.32	1.40
40	1.20	1.31	0.77	1.21	1.23	1.27	1.28	1.30	1.20	1.20	1.21	1.24	1.30
45	1.14	1.24	0.73	1.15	1.17	1.21	1.22	1.24	1.14	1.14	1.15	1.18	1.24
50	1.08	1.18	0.69	1.09	1.11	1.14	1.15	1.17	1.08	1.08	1.09	1.11	1.17

Table 5.5: Confidence length of 95% CIs when skewness=8

n	t	med	mad	Trimmed t CI (%)					Modified Trimmed t CI (%)				
				5%	10%	20%	30%	45%	5%	10%	20%	30%	45%
5	4.70	5.35	3.42	4.70	4.70	5.18	5.18	5.35	4.70	4.70	4.70	5.14	5.35
10	3.51	3.82	2.21	3.51	3.67	3.77	3.77	3.82	3.51	3.51	3.70	3.74	3.82
15	3.01	3.22	1.73	3.01	3.09	3.19	3.19	3.22	3.01	3.04	3.15	3.18	3.22
20	2.69	2.86	1.47	2.75	2.79	2.84	2.85	2.86	2.69	2.72	2.82	2.85	2.86
25	2.55	2.69	1.33	2.59	2.63	2.68	2.68	2.69	2.55	2.57	2.66	2.68	2.69
30	2.28	2.40	1.16	2.31	2.36	2.39	2.40	2.40	2.29	2.30	2.38	2.39	2.40
35	2.21	2.32	1.10	2.24	2.28	2.31	2.32	2.32	2.21	2.23	2.30	2.32	2.32
40	2.12	2.22	1.04	2.16	2.19	2.22	2.22	2.22	2.12	2.13	2.21	2.22	2.22
45	2.02	2.11	0.97	2.05	2.08	2.10	2.11	2.11	2.02	2.03	2.10	2.11	2.11
50	1.92	2.00	0.92	1.94	1.98	2.00	2.00	2.00	1.92	1.93	1.99	2.00	2.00

Table 5.6: Minimum (min) and maximum (max) confidence length of 95% confidence interval for varying values of skewness and % trimming

Skewness=0.5													
n	t	med	mad	Trimmed t CI (%)					Modified Trimmed t CI (%)				
				5%	10%	20%	30%	45%	5%	10%	20%	30%	45%
min	0.14	0.14	0.11	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14
max	0.58	0.61	0.44	0.58	0.58	0.59	0.59	0.61	0.58	0.58	0.58	0.58	0.61
Skewness=1													
min	0.28	0.29	0.22	0.28	0.28	0.28	0.28	0.29	0.28	0.28	0.28	0.28	0.29
max	1.14	1.21	0.87	1.14	1.14	1.17	1.17	1.21	1.14	1.14	1.14	1.15	1.21
Skewness=2													
min	0.56	0.58	0.41	0.56	0.57	0.57	0.58	0.58	0.56	0.56	0.56	0.56	0.58
max	2.14	2.31	1.62	2.14	2.14	2.21	2.21	2.31	2.14	2.14	2.14	2.18	2.30
Skewness=4													
min	1.08	1.18	0.69	1.09	1.11	1.14	1.15	1.17	1.08	1.08	1.09	1.11	1.17
max	3.59	4.03	2.67	3.59	3.59	3.83	3.83	4.03	3.59	3.59	3.59	3.77	4.03
Skewness=8													
min	1.92	2.00	0.92	1.94	1.98	2.00	2.00	2.00	1.92	1.93	1.99	2.00	2.00
max	4.70	5.35	3.42	4.70	4.70	5.18	5.18	5.35	4.70	4.70	4.70	5.14	5.35

The simulation results suggest that when the skewness is 0.5 (Table 4.1), all methods, except the med t CI, perform reasonably well with coverage probability equal to the nominal level of 0.95, or within 1% of the nominal level of 0.95. As reported in Table 4.1, the mad t CI has the lowest coverage probability with coverage probability of 95% CI ranging from 0.88 to 0.90.

As skewness increases from 0.50 to 8, severe underestimation is observed for mad t CI, while it has the shortest observed confidence length in all simulation cases have been studied (Tables 5.1-5.6). The min coverage probability of all confidence interval decreases with increases skewness (Table 4.6). In all simulation cases, modified trimmed t CI has the highest minimum or highest maximum coverage probability or coverage probability of CIs similar to Student's t , trimmed t or median t . Overall, modified trimmed t CI retains the efficiency of Student's t and robustness of median t or trimmed t as is evident in the estimated coverage probability.

It is also noted that the coverage probability is sensitive to (i) the sample size and (ii) level of skewness. As sample size increases, the coverage probability increase for higher skewness. As skewness increases, the coverage probability decreases. For a fixed value of skewness, modified trimmed t CI has the highest coverage probability or coverage probability equal to the med t CI. With higher % trimming, coverage probability of trimmed t CI approaches the coverage probability of the med t CI. Clearly, these results suggest that trimmed t and modified t CI retains the efficiency of Student's t CI and the robustness of median t CI.

On the other hand, if confidence length is concerned in the compromise of the coverage probability, then mad t CI has the shortest confidence length in all simulation cases studied. Confidence length is sensitive to sample size in that with increase sample size, confidence length decreases for all CIs. Confidence length is also sensitive to skewness in that the confidence length increases as the skewness increases.

6. Concluding Remarks

If population distribution is skewed, then the modified trimmed t CI proposed in this article retains the highest coverage probability or equally highest coverage probability with med t CI. With increasing % trimming, the performance of the trimmed t is as good as the median t CI. Mad t CI has the lowest coverage probability. With lower % trimmed, trimmed and modified trimmed t CI are identical or close to the Student's t CI. The coverage probability of all CIs decreases with the increase in skewness, and for highly skewed distribution coverage probability increases with the increase in the sample size. At a fixed value of the skewness, modified trimmed t CI has the highest coverage probability or coverage probability equal to the med t CI. With higher % trimming, the performance of the trimmed t is comparable with the med t CI. In all circumstances, the proposed modified

trimmed t CI performs satisfactorily. Therefore, given any indication of skewness, the modified trimmed t CI should be considered positively for estimating the CI of the true population mean.

References

- Hayden, R.W. (2005). A Dataset that is 44% Outliers. *Journal of Statistics Education*, 13 (1).
- Johnson, N.J. (1978). Modified t Tests and Confidence Intervals for Asymmetrical Populations. *Journal of the American Statistical Association*, 73, pp. 536-544.
- Kibria, B.M.G. (2006). Modified Confidence Intervals for the Mean of the Asymmetric Distribution. *Pakistan Journal of Statistics*, 22(2), pp. 111-123.
- Kleijnen J.P.C., Kloppenburg, G.L.J. and Meeuwssen, F.L. (1986). Testing the mean of asymmetric population: Johnson's modified t test revisited. *Communications in Statistics- Simulation and Computation*, 15, 715-732.
- Meeden, G. (1999). Interval Estimators for the Population Mean for Skewed Distributions with a Small Sample Size. *Journal of Applied Statistics*, 26(1), 81-96.
- R version 3.3.2 (2016-10-31). *The R Foundation for Statistical Computing*.
- Shi, W. and Kibria, B.M.G. (2007). On some confidence intervals for estimating the mean of a skewed population. *Int. J. Math. Educ. Sci. Technol.* 38(3), pp. 412-421.
- Student (1908). The probable error of a mean. *Biometrika* 6 (1): 1-25.
- Willink, R. (2005). A Confidence Interval and Test for the Mean of an Asymmetric Distribution. *Communications in Statistics- Theory and Methods*, 34, 753-766.
- Wrona, R.M. (1979). A clinical epidemiologic study of hyperphenylalaninemia. *American Journal of Public Health* July,69(7) pp. 673-679.