8-2021

Some t-Type Confidence Intervals

Daina McKinney
daina.mckinney@siu.edu

SOME t-TYPE CONFIDENCE INTERVALS

by

Daina McKinney

B.A., Southern Illinois University Carbondale, 2019

A Research Paper
Submitted in Partial Fulfillment of the Requirements for the
Master of Science

Department of Mathematics
in the Graduate School
Southern Illinois University Carbondale
July, 2021

RESEARCH PAPER APPROVAL

SOME t-TYPE CONFIDENCE INTERVALS

by

Daina McKinney

A Research Paper Submitted in Partial

Fulfillment of the Requirements

for the Degree of

Master of Science

in the field of Mathematics

Approved by:

David J. Olive

Wesley Calvert

Kwangho Choiy

Graduate School
Southern Illinois University Carbondale
June 30, 2021

AN ABSTRACT OF THE RESEARCH PAPER OF

Daina McKinney, for the Master of Science degree in Mathematics, presented on June 30, 2021, at Southern Illinois University Carbondale.

TITLE: SOME t-TYPE CONFIDENCE INTERVALS

MAJOR PROFESSOR: Dr. David J. Olive

This paper compares several competitors of the t confidence interval for the population mean μ given iid data $Y_1, ..., Y_n$.

ACKNOWLEDGMENTS

I would like to take this opportunity to thank my research advisor, Dr. David Olive for overseeing my Master's project and Dr. Wesley Calvert and Dr, Kwangho Choiy for sitting on my committee. I would also like to thank the professors of Southern Illinois University for their instruction and care over the past four years. Finally, I want to say to my family (especially my husband) thank you for all of your support and encouragement. You have pushed me to succeed and I could not have done it without you!

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 1

INTRODUCTION

Definition 1. Let the data $Y_1, ..., Y_n$ have joint probability density function (pdf) or probability mass function (pmf) $f(\boldsymbol{y}|\theta)$ with parameter space Θ and support $\mathcal{Y}$. Let $L_n = L_n(\boldsymbol{Y})$ and $U_n = U_n(\boldsymbol{Y})$ be statistics such that $L_n(\boldsymbol{y}) \leq U_n(\boldsymbol{y})$, $\forall \boldsymbol{y} \in \mathcal{Y}$. Then $[L_n, U_n]$ is a large sample 100 $(1-\alpha)$ % CI for θ if

$$P_\theta(L_n(\boldsymbol{Y}) \leq \theta \leq U_n(\boldsymbol{Y}))$$

is eventually bounded below by $1-\alpha$ for all $\theta \in \Theta$ as $n \to \infty$.

The following statistics will be useful for constructing large sample confidence intervals.

Definition 2. Let the $Y_1, ..., Y_n$ be the data.

a) The *sample mean*

$$\overline{Y} = \frac{\sum_{i=1}^n Y_i}{n}. \quad (1.1)$$

b) The *sample variance*

$$S^2 \equiv S_n^2 = \frac{\sum_{i=1}^n (Y_i - \overline{Y})^2}{n-1} = \frac{\sum_{i=1}^n Y_i^2 - n(\overline{Y})^2}{n-1}. \quad (1.2)$$

c) The *sample standard deviation* $S \equiv S_n = \sqrt{S_n^2}$.

d) If the data $Y_1, ..., Y_n$ is arranged in ascending order from smallest to largest and written as $Y_{(1)} \leq \cdots \leq Y_{(n)}$, then $Y_{(i)}$ is the ith order statistic and the $Y_{(i)}$'s are called the *order statistics*.

To find a large sample CI for $E(Y) = \mu$, let z_α be the 100αth percentile of the standard normal $N(0,1)$ distribution. Hence $P(Z \leq z_\alpha) = \alpha$ if $Z \sim N(0,1)$. Similarly, let $t_{n-1,\alpha}$ satisfy $P(X \leq t_{n-1,\alpha}) = \alpha$ if $X \sim t_{n-1}$, a t distribution with $n-1$ degrees of freedom. Using $z_\alpha = -z_{1-\alpha}$ for $0 < \alpha < 0.5$, note that for large n,

$$1 - \alpha \approx P(-z_{1-\alpha/2} \leq \frac{\overline{Y} - \mu}{S/\sqrt{n}} \leq z_{1-\alpha/2}) =$$

$$P(-z_{1-\alpha/2} \;\; S/\sqrt{n} \le \overline{Y} - \mu \le z_{1-\alpha/2} \;\; S/\sqrt{n}) =$$

$$P(-\overline{Y} - z_{1-\alpha/2} \;\; S/\sqrt{n} \le -\mu \le -\overline{Y} + z_{1-\alpha/2} \;\; S/\sqrt{n}) =$$

$$P(\overline{Y} - z_{1-\alpha/2} \;\; S/\sqrt{n} \le \mu \le \overline{Y} + z_{1-\alpha/2} \;\; S/\sqrt{n}).$$

Then

$$\overline{Y} \pm z_{1-\alpha/2} \;\; S/\sqrt{n}$$

is a large sample $100(1-\alpha)\%$ CI for μ.

Since $t_{n-1,1-\alpha/2} > z_{1-\alpha/2}$ but $t_{n-1,1-\alpha/2} \to z_{1-\alpha/2}$ as $n \to \infty$,

$$[\ \overline{Y} - t_{n-1,1-\alpha/2} \;\; S/\sqrt{n}, \;\; \overline{Y} + t_{n-1,1-\alpha/2} \;\; S/\sqrt{n}\ ] \tag{1.3}$$

is also a large sample $100(1-\alpha)\%$ CI for μ. This t CI is one of the most widely used confidence intervals in Statistics. Replacing $z_{1-\alpha/2}$ by $t_{n-1,1-\alpha/2}$ makes the CI longer and hence less likely to be liberal: actual coverage is $< 1-\alpha$ for a given n.

CHAPTER 2

SOME COMPETITORS OF THE *t* INTERVAL

Hesterberg (2014) claims that the central limit theorem and the t CI (1.3) need $n \geq 5000$ for moderately skewed distributions such as the Exponential(μ) distribution with (population) mean $E(Y) = \mu$ and variance $V(Y) = \sigma^2 = \mu^2$. Moore (2013, pp. 298-299) shows that the distribution for $\overline{Y}$ is approximately normal for $n = 25$ for the Exponential(μ) distribution.

Hesterberg's (2014) claim is based on using a t CI where the length of the CI is "short" (use $c = 0.1\mu$ below) with high probability. To help see why large n might be needed, assume $\overline{Y}_n$ is approximately normal if n is large enough such that

$$P(\mu - c \leq \overline{Y}_n \leq \mu + c) \geq \Phi(2) - \Phi(-2) = 0.9544$$

where $\Phi(x)$ is the cumulative distribution function of the N(0,1) distribution. Then

$$P\left(\frac{-c\sqrt{n}}{\sigma} \leq \frac{\overline{Y}_n - \mu}{\sigma/\sqrt{n}} \leq \frac{\sqrt{n}}{\sigma}\right) \approx \Phi\left(\frac{c\sqrt{n}}{\sigma}\right) - \Phi\left(\frac{-c\sqrt{n}}{\sigma}\right) \approx \Phi(2) - \Phi(-2),$$

or $c\sqrt{n}/\sigma \geq 2$ or

$$n \geq \frac{4\sigma^2}{c^2}.$$

Hesterberg (2014) takes $c = 0.1\mu$, hence $n \geq 400$ for the Exponential(μ) distribution. Note that if $c = 0.5\mu$, then $n \geq 16$ although we likely need $n \geq 25$ for the normal approximation for $\overline{Y}_n$ to be good for the Exponential distribution. The same calculation $n \geq 400$ also holds for the $N(\mu, \mu^2)$ distribution, where the t CI coverage is exact. The t interval estimates σ with S, and hence needs larger sample size n than the CLT.

For the t interval, S is also used, and Hesterberg (2014) suggests that for

$$\frac{\overline{Y} - \mu}{S/\sqrt{n}}$$

to have an approximate t_{n-1} distribution, we need

$$n \geq \left(\frac{\gamma}{6} \frac{10}{\alpha}(2z^2_{\alpha/2} + 1)\phi(z_{\alpha/2})\right)^2$$

where $\phi(x)$ is the N(0,1) pdf and

$$\gamma = \frac{E[(Y-\mu)^3]}{\sigma^3}$$

is the skewness of the distribution. It is not clear how this n depends on $c = k\mu$ with $k \in [0.1, 0.5]$.

Hesterberg (2014) gives the following two competitors of the t interval: the skewness adjusted t inteval is

$$[\ \overline{Y} + \frac{S}{\sqrt{n}}[\hat{\kappa}(1 + 2t^2_{n-1,1-\alpha/2}) - t_{n-1,1-\alpha/2}],\ \overline{Y} + \frac{S}{\sqrt{n}}[\hat{\kappa}(1 + 2t^2_{n-1,1-\alpha/2}) + t_{n-1,1-\alpha/2}]\ ], \quad (2.1)$$

and the asymptotic percentile t CI is

$$[\ \overline{Y} + \frac{S}{\sqrt{n}}[\hat{\kappa}(t_{n-1,1-\alpha/2} - 1)^2 - t_{n-1,1-\alpha/2}],\ \overline{Y} + \frac{S}{\sqrt{n}}[\hat{\kappa}(t_{n-1,1-\alpha/2} - 1)^2 + t_{n-1,1-\alpha/2}] \quad (2.2)$$

where

$$\hat{\kappa} = \frac{\hat{\gamma}}{6\sqrt{n}} \ \text{ with } \ \hat{\gamma} = \frac{1}{\mathrm{nS^3}} \sum_{\mathrm{i=1}}^{\mathrm{n}} (\mathrm{Y_i} - \overline{\mathrm{Y}})^3.$$

The Johnson (1978) CI is

$$[\ \overline{Y} + \frac{\hat{\mu}_3}{6S^2n} - t_{n-1,1-\alpha/2}\ \ S/\sqrt{n},\ \ \overline{Y} + \frac{\hat{\mu}_3}{6S^2n} + t_{n-1,1-\alpha/2}\ \ S/\sqrt{n}\ ] \quad (2.3)$$

where $\mu_3 = E[(Y-\mu)^3]$ and

$$\hat{\mu}_3 = S^3\hat{\gamma} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \overline{Y})^3.$$

CI (2.3) approximates the percentile CI (2.5). CIs (2.2) and (2.3) are the same if $t_{n-1,1-\alpha/2} = 2$, so are similar for 95% CIs if n is large. The CIs (2.1)-(2.3) need the distribution of the iid Y_i to have third moments. The CIs (1.1), (2.4), and (2.5) need second moments. We will also consider two CIs based on the nonparametric bootstrap which draws n cases with replacement from the data set $Y_1, ..., Y_n$. Then T_1^* is the statistic T_n computed from the sample. This process is repeated B times to produce the bootstrap sample $T_1^*, ..., T_B^*$.

Definition 3. Suppose that data $\boldsymbol{x}_1, ..., \boldsymbol{x}_n$ has been collected and observed. Often the data is a random sample (iid) from a distribution with cdf F. The *empirical distribution* is a discrete distribution where the $\boldsymbol{x}_i$ are the possible values, and each value is equally likely. If $\boldsymbol{w}$ is a random variable having the empirical distribution, then $p_i = P(\boldsymbol{w} = \boldsymbol{x}_i) = 1/n$ for $i = 1, ..., n$. The *cdf of the empirical distribution* is denoted by F_n.

Example 1. Let $\boldsymbol{w}$ be a random variable having the empirical distribution given by Definition 3. Show that $E(\boldsymbol{w}) = \overline{\boldsymbol{x}} \equiv \overline{\boldsymbol{x}}_n$ and $\text{Cov}(\boldsymbol{w}) = \dfrac{n-1}{n}\boldsymbol{S} \equiv \dfrac{n-1}{n}\boldsymbol{S}_n$.

Solution: Recall that for a discrete random vector, the population expected value $E(\boldsymbol{w}) = \sum \boldsymbol{x}_i p_i$ where $\boldsymbol{x}_i$ are the values that $\boldsymbol{w}$ takes with positive probability p_i. Similarly, the population covariance matrix

$$\text{Cov}(\boldsymbol{w}) = E[(\boldsymbol{w} - E(\boldsymbol{w}))(\boldsymbol{w} - E(\boldsymbol{w}))^T] = \sum (\boldsymbol{x}_i - E(\boldsymbol{w}))(\boldsymbol{x}_i - E(\boldsymbol{w}))^T p_i.$$

Hence

$$E(\boldsymbol{w}) = \sum_{i=1}^{n} \boldsymbol{x}_i \frac{1}{n} = \overline{\boldsymbol{x}},$$

and

$$\text{Cov}(\boldsymbol{w}) = \sum_{i=1}^{n} (\boldsymbol{x}_i - \overline{\boldsymbol{x}})(\boldsymbol{x}_i - \overline{\boldsymbol{x}})^T \frac{1}{n} = \frac{n-1}{n}\boldsymbol{S}. \ \square$$

Example 2. Suppose the data is 1, 2, 3, 4, 5, 6, 7. Then $n = 7$ and the sample median T_n is 4. Using R, we drew $B = 2$ bootstrap samples (samples of size n drawn with replacement from the original data) and computed the sample median $T^*_{1,n} = 3$ and $T^*_{2,n} = 4$.

```
b1 <- sample(1:7,replace=T)
b1
[1] 3 2 3 2 5 2 6
median(b1)
[1] 3
b2 <- sample(1:7,replace=T)
```

```
b2
[1] 3 5 3 4 3 5 7
median(b2)
[1] 4
```

Definition 4. The bootstrap percentile method large sample $100(1-\delta)\%$ confidence interval for θ is an interval $[T^*_{(k_L)}, T^*_{(K_U)}]$ containing $\approx \lceil B(1-\delta) \rceil$ of the T^*_i. Let $k_1 = \lceil B\delta/2 \rceil$ and $k_2 = \lceil B(1-\delta/2) \rceil$. A common choice is

$$[T^*_{(k_1)}, T^*_{(k_2)}]. \tag{2.4}$$

Definition 5. The large sample $100(1-\delta)\%$ *shorth*(c) *CI*

$$[T^*_{(s)}, T^*_{(s+c-1)}]. \tag{2.5}$$

uses the interval $[T^*_{(1)}, T^*_{(c)}]$, $[T^*_{(2)}, T^*_{(c+1)}]$, ..., $[T^*_{(B-c+1)}, T^*_{(B)}]$ of shortest length. Here

$$c = \min(B, \lceil B[1 - \delta + 1.12\sqrt{\delta/B}\ ]\ \rceil).$$

Example 3. Given below were votes for preseason 1A basketball poll from Nov. 22, 2011 WSIL News where the 778 was a typo: the actual value was 78. As shown below, finding shorth(3) from the ordered data is simple. If the outlier was corrected, shorth(3) = [76,78].

```
111   89   778   78   76


order data: 76 78 89 111 778


                  13 = 89 - 76
```

```
                    33 = 111 - 78


                         689 = 778 - 89

shorth(3) = [76,89]
```

There is considerable theory for the percentile method. Pelawa Watagoda and Olive (2021) note that if $\sqrt{n}(T_n - \theta) \xrightarrow{D} U$ and $\sqrt{n}(T_i^* - T_n) \xrightarrow{D} U$ where U has a unimodal probability density function symmetric about zero, then the shorth confidence interval (2.5), and the "usual" percentile method confidence interval (2.4) are asymptotically equivalent (use the central proportion of the bootstrap sample, asymptotically). These conditions hold for the sample mean. The CI (2.4) applies a nonparametric prediction interval to the bootstrap sample. The shorth CI is the Frey (2013) prediction interval applied to the bootstrap sample, and can be regarded as the shortest percentile method confidence interval, asymptotically. Hence the shorth confidence interval is a practical implementation of the Hall (1988) shortest bootstrap confidence interval based on all possible bootstrap samples.

Hesterberg (2014, p. 51) noted that the t interval is more accurate than the percentile interval for $n \leq 34$. Note that CIs (1.3)-(2.3) all have the same length. From Example 1, the variability of the sample mean computed from the nonparametric bootstrap is about $(n-1)/n$ times that of the sample mean computed from iid data.

Consider distributions with second moments. See Olive (2014, ch. 10) for some brand name distributions. The t CI tends to work well even if $n = 30$ for moderately right skewed distribution such as the half normal or exponential distribution. The t CI needs large n for highly skewed distributions. One way to get a highly skewed distribution is to let $Y = e^X$ or $Y = X^k$ for large integer k where X is a moderately right skewed distribution. Hence the lognormal, Pareto, and some Weibull distribution are highly skewed. If $Y = \sum_{i=1}^n X_i$ is not approximatley normal, then the CLT does not work well. Some infinitely divisible distributions have this property. For example, if $Y \sim$ Poisson(μ), then $Y = \sum_{i=1}^n X_i$ where

$X_i \sim$ Poisson(μ/n). The normal approximation starts to be reasonable for $\mu \geq 9$. Hence if $m\mu = 1$ for positive integer n, then need $n \geq 9m$. Similarly, the normal appoximation starts to be reasonable for $Y \sim$ binomial$(m, p = 0.5)$ for $m \geq 10$, but the normal approximation is bad if mp is near 0 or m.

CHAPTER 3

SIMULATIONS

A large sample CI gives "reasonable values" of the parameter $\theta = \mu$. We will say that the simulated CI is best if for observed coverage ≥ 0.94, the CI has the shortest simulated average length. A small simulation was done using 5000 runs. So an observed coverage in [0.94, 0.96] gives no reason to doubt that the CI has the nominal coverage of 0.95. The errors e_i were iid from 9 distributions shown in the tables: i) N(0,1), ii) t_3, iii) EXP(1), iv) uniform$(-1, 1)$, v) the mixture 0.1 N(0,1) + 0.9 N(0,100), vi) Poisson(9), vii) binomial(10,0.5), viii) lognormal(0,1), and ix) Pareto$(\sigma = 1, \lambda = 0.25)$. The t_3 distribution has second moments but not third. The asymptotic average scaled length of each 95% CI (except possibly distribution ii) for CIs that estimate the third moment), where the scaling multiplies the length by $\sqrt{n}$, is $2(1.96)\sigma = 3.92\sigma$. These scaled lengths are i) 3.92, ii) 6.7896, iii) 3.92, iv) 2.2632, v) 12.9419, vi) 11.76, vii) 6.1981, viii) 8.4719, and ix) 1.8479. In the simulations, only the highly skewed lognormal and Pareto distributions needed large n to have coverage and scaled length near 0.95 and the asymptotic scaled length.

The *R* function `tcisim` uses the iid error distributions 1: N(0,1), 2: t_3, 3: exponential(1), 4: uniform(-1,1), 5: 0.1 N(0,1) + 0.9 N(0,100), 6: Poisson(9), 7: binomial(10,0.5), 8: lognormal(0,1), and 9: for Pareto(1,1/4) corresponding to type. The function makes the t CI (1.3), shorth CI (2.5), percentile CI (2.4), Johnson CI (2.3), skewness adjusted t CI (2.1), and asymptotic percentile t CI (2.2). The bootstrap CIs (2.4) and (2.5) used $B = 1000$ bootstrap samples. The simulations used 5000 runs and cov is the proportion of times the CI contained the mean μ. The average scaled CI lengths were also given where the CI length was multiplied by $\sqrt{n}$.

```
source("http://parker.ad.siu.edu/Olive/sipack.txt")
args(tcisim)
function (n = 100, BB = 1000, nruns = 100, type = 1, dd = 3,
```

Table 3.1. CI simulation, n=30, B=1000

e	t	shorth	perc	J	adj	asy
N,cov	0.9532	0.9396	0.9350	0.9532	0.9518	0.9534
len	4.0681	3.9000	3.8111	4.0681	4.0681	4.0681
t,cov	0.955	0.9392	0.9236	0.9528	0.9016	0.9526
len	6.5236	6.2142	6.0989	6.5236	6.5236	6.5236
E,cov	0.9234	0.9158	0.9154	0.9258	0.9308	0.9258
len	3.9757	3.7737	3.7084	3.9757	3.9757	3.9757
U,cov	0.9516	0.9438	0.9412	0.9534	0.9622	0.9534
len	2.3560	2.2561	2.2051	2.3560	2.3560	2.3560
MIX,cov	0.9752	0.9388	0.9024	0.968	0.7518	0.9674
len	12.3452	11.56307	11.4383	12.3452	12.3452	12.3452
Pois,cov	0.9506	0.9444	0.9396	0.9506	0.9498	0.9506
len	12.1859	11.6454	11.413	12.1859	12.1859	12.1859
Bin,cov	0.9478	0.944	0.9392	0.948	0.9488	0.948
len	6.4097	6.0989	6.0030	6.4097	6.4097	6.4097
LN,cov	0.8898	0.8826	0.8806	0.895	0.881	0.8952
len	7.8159	7.2671	7.2158	7.8159	7.8159	7.8159
Par,cov	0.8944	0.884	0.8872	0.8984	0.8918	0.8992
len	1.7140	1.5969	1.5826	1.7140	1.7140	1.7140

```
    eps = 0.1, shift = 9, alph = 0.05)


tcisim(n=30,type=1,nruns=5000)
#N(0,1) asymptotic scaled length 3.92
$tcicov
[1] 0.9532
$tslen
[1] 4.068066
$shorthcov
[1] 0.9396
$shslen
[1] 3.900022
$perccov
[1] 0.935
$percslen
3.811119
$jcicov
[1] 0.9532
$jslen
[1] 4.068066
$adjcov
[1] 0.9518
$adjslen
[1] 4.068066
$asycov
[1] 0.9534
```

```
$asyslen
[1] 4.068066

tcisim(n=30,type=2,nruns=5000)
#t3 asymptotic scaled length 6.7896
$tcicov
[1] 0.955
$tslen
[1] 6.523621
$shorthcov
[1] 0.9392
$shslen
[1] 6.214155
$perccov
[1] 0.9236
$percslen
6.098922
$jcicov
[1] 0.9528
$jslen
[1] 6.523621
$adjcov
[1] 0.9016
$adjslen
[1] 6.523621
$asycov
[1] 0.9526
```

```
$asyslen
[1] 6.523621


tcisim(n=30,type=3,nruns=5000)
#Exp(1) asymptotic scaled length 3.92
$tcicov
[1] 0.9234
$tslen
[1] 3.975708
$shorthcov
[1] 0.9158
$shslen
[1] 3.773712
$perccov
[1] 0.9154
$percslen
3.708418
$jcicov
[1] 0.9258
$jslen
[1] 3.975708
$adjcov
[1] 0.9308
$adjslen
[1] 3.975708
$asycov
[1] 0.9258
```

```
$asyslen
[1] 3.975708

tcisim(n=30,type=4,nruns=5000)
#U(-1,1) asymptotic scaled length 2.2632
$tcicov
[1] 0.9516
$tslen
[1] 2.356043
$shorthcov
[1] 0.9438
$shslen
[1] 2.256136
$perccov
[1] 0.9412
$percslen
2.205131
$jcicov
[1] 0.9534
$jslen
[1] 2.356043
$adjcov
[1] 0.9622
$adjslen
[1] 2.356043
$asycov
[1] 0.9534
```

```
$asyslen
[1] 2.356043


tcisim(n=30,type=5,nruns=5000)
#mixture asymptotic scaled length 12.9419
$tcicov
[1] 0.9752
$tslen
[1] 12.34523
$shorthcov
[1] 0.9388
$shslen
[1] 11.56307
$perccov
[1] 0.9024
$percslen
11.4383
$jcicov
[1] 0.968
$jslen
[1] 12.34523
$adjcov
[1] 0.7518
$adjslen
[1] 12.34523
$asycov
[1] 0.9674
```

```
$asyslen
[1] 12.34523

tcisim(n=30,type=6,nruns=5000)
#Poisson(9) asymptotic scaled length 11.76
$tcicov
[1] 0.9506
$tslen
[1] 12.18588
$shorthcov
[1] 0.9444
$shslen
[1] 11.64542
$perccov
[1] 0.9396
$percslen
11.413
$jcicov
[1] 0.9506
$jslen
[1] 12.18588
$adjcov
[1] 0.9498
$adjslen
[1] 12.18588
$asycov
[1] 0.9506
```

```
$asyslen
[1] 12.18588


tcisim(n=30,type=7,nruns=5000)
#bin(10,0.5) asymptotic scaled length 6.1981
$tcicov
[1] 0.9478
$tslen
[1] 6.409742
$shorthcov
[1] 0.944
$shslen
[1] 6.098854
$perccov
[1] 0.9392
$percslen
[1] 6.003031
$jcicov
[1] 0.948
$jslen
[1] 6.409742
$adjcov
[1] 0.9488
$adjslen
[1] 6.409742
$asycov
```

```
[1] 0.948

$asyslen

[1] 6.409742


tcisim(n=30,type=8,nruns=5000) #need n large

#lognormal(0,1) asymptotic scaled length 8.4719

$tcicov

[1] 0.8898

$tslen

[1] 7.815919

$shorthcov

[1] 0.8826

$shslen

[1] 7.26707

$perccov

[1] 0.8806

$percslen

7.215783

$jcicov

[1] 0.895

$jslen

[1] 7.815919

$adjcov

[1] 0.881

$adjslen

[1] 7.815919

$asycov
```

```
[1] 0.8952

$asyslen

[1] 7.815919


tcisim(n=30,type=9,nruns=5000) #need n large

#Pareto(1,0.25) asymptotic scaled length 1.8479

$tcicov

[1] 0.8944

$tslen

[1] 1.713966

$shorthcov

[1] 0.884

$shslen

[1] 1.596903

$perccov

[1] 0.8872

$percslen

   97.5%

1.582632

$jcicov

[1] 0.8984

$jslen

[1] 1.713966

$adjcov

[1] 0.8918

$adjslen

[1] 1.713966
```

```
$asycov
[1] 0.8992
$asyslen
[1] 1.713966


source("http://parker.ad.siu.edu/Olive/sipack.txt")
args(tcisim)
function (n = 100, BB = 1000, nruns = 100, type = 1, dd = 3,
    eps = 0.1, shift = 9, alph = 0.05)


tcisim(n=100,type=1,nruns=5000)
#N(0,1) asymptotic scaled length 3.92
$tcicov
[1] 0.9478
$tslen
[1] 3.958718
$shorthcov
[1] 0.9498
$shslen
[1] 3.966194
$perccov
[1] 0.943
$percslen
3.873937
$jcicov
```

Table 3.2. CI simulation, n=100, B=1000

e	t	shorth	perc	J	adj	asy
N,cov	0.9478	0.9498	0.943	0.9476	0.9476	0.9476
len	3.9587	3.9662	3.8739	3.9587	3.9587	3.9587
t,cov	0.9568	0.9518	0.946	0.9548	0.9264	0.955
len	6.5654	6.5566	6.4151	6.5654	6.5654	6.5654
E,cov	0.9394	0.9384	0.9364	0.9402	0.9406	0.9398
len	3.9209	3.9134	3.8310	3.9209	3.9209	3.9209
U,cov	0.9522	0.9506	0.9454	0.9524	0.9554	0.9524
len	2.2898	2.2910	2.2384	2.2898	2.2898	2.2898
MIX,cov	0.9526	0.9358	0.9142	0.9446	0.858	0.9452
len	12.6792	12.7009	12.4333	12.6792	12.6792	12.6792
Pois,cov	0.9506	0.9508	0.9476	0.9506	0.9506	0.9506
len	11.8621	11.8659	11.6069	11.8621	11.8621	11.8621
Bin,cov	0.9496	0.9552	0.9482	0.9498	0.9506	0.9498
len	6.2616	6.2517	6.1261	6.2616	6.2616	6.2616
LN,cov	0.9136	0.9138	0.9144	0.9164	0.907	0.9164
len	8.0963	7.9846	7.8705	8.0963	8.0963	8.0963
Par,cov	0.9218	0.9228	0.9214	0.9234	0.9118	0.9234
len	1.7723	1.7490	1.7228	1.7723	1.7723	1.7723

```
[1] 0.9476

$jslen

[1] 3.958718

$adjcov

[1] 0.9474

$adjslen

[1] 3.958718

$asycov

[1] 0.9475

$asyslen

[1] 3.958718


tcisim(n=100,type=2,nruns=5000)

#t3 asymptotic scaled length 6.7896

$tcicov

[1] 0.9568

$tslen

[1] 6.565404

$shorthcov

[1] 0.9518

$shslen

[1] 6.556578

$perccov

[1] 0.946

$percslen

6.415138

$jcicov
```

```
[1] 0.9548

$jslen

[1] 6.565404

$adjcov

[1] 0.9264

$adjslen

[1] 6.565404

$asycov

[1] 0.955

$asyslen

[1] 6.565404


tcisim(n=100,type=3,nruns=5000)

#Exp(1) asymptotic scaled length 3.92

$tcicov

[1] 0.9394

$tslen

[1] 3.92086

$shorthcov

[1] 0.9384

$shslen

[1] 3.913403

$perccov

[1] 0.9364

$percslen

3.830999

$jcicov
```

```
[1] 0.9402

$jslen

[1] 3.92086

$adjcov

[1] 0.9406

$adjslen

[1] 3.92086

$asycov

[1] 0.9398

$asyslen

[1] 3.92086


tcisim(n=100,type=4,nruns=5000)

#U(-1,1) asymptotic scaled length 2.2632

$tcicov

[1] 0.9522

$tslen

[1] 2.289766

$shorthcov

[1] 0.9506

$shslen

[1] 2.290954

$perccov

[1] 0.9454

$percslen

2.238419

$jcicov
```

```
[1] 0.9524

$jslen

[1] 2.289766

$adjcov

[1] 0.9554

$adjslen

[1] 2.289766

$asycov

[1] 0.9524

$asyslen

[1] 2.289766


tcisim(n=100,type=5,nruns=5000)

#mixture asymptotic scaled length 12.9419

$tcicov

[1] 0.9526

$tslen

[1] 12.67915

$shorthcov

[1] 0.9358

$shslen

[1] 12.70092

$perccov

[1] 0.9142

$percslen

12.43325

$jcicov
```

```
[1] 0.9446

$jslen

[1] 12.67915

$adjcov

[1] 0.858

$adjslen

[1] 12.67915

$asycov

[1] 0.9452

$asyslen

[1] 12.67915


tcisim(n=100,type=6,nruns=5000)

#Poisson(9) asymptotic scaled length 11.76

$tcicov

[1] 0.9506

$tslen

[1] 11.86208

$shorthcov

[1] 0.9508

$shslen

[1] 11.86692

$perccov

[1] 0.9475

$percslen

11.60694

$jcicov
```

```
[1] 0.9506

$jslen

[1] 11.86208

$adjcov

[1] 0.9506

$adjslen

[1] 11.86208

$asycov

[1] 0.9506

$asyslen

[1] 11.86208



tcisim(n=100,type=7,nruns=5000)

#bin(10,0.5) asymptotic scaled length 6.1981

$tcicov

[1] 0.9496

$tslen

[1] 6.261562

$shorthcov

[1] 0.9552

$shslen

[1] 6.25168

$perccov

[1] 0.9482

$percslen

[1] 6.126082
```

```
$jcicov
[1] 0.9498
$jslen
[1] 6.261562
$adjcov
[1] 0.9506
$adjslen
[1] 6.261562
$asycov
[1] 0.9498
$asyslen
[1] 6.261562

tcisim(n=100,type=8,nruns=5000)  #need n large
#lognormal(0,1) asymptotic scaled length 8.4719
$tcicov
[1] 0.9136
$tslen
[1] 8.096294
$shorthcov
[1] 0.9138
$shslen
[1] 7.984596
$perccov
[1] 0.9144
$percslen
7.870549
```

```
$jcicov
[1] 0.9164
$jslen
[1] 8.096294
$adjcov
[1] 0.907
$adjslen
[1] 8.096294
$asycov
[1] 0.9164
$asyslen
[1] 8.096294


tcisim(n=100,type=9,nruns=5000)  #need n large
#Pareto(1,0.25) asymptotic scaled length 1.8479
$tcicov
[1] 0.9218
$tslen
[1] 1.772316
$shorthcov
[1] 0.9228
$shslen
[1] 1.748951
$perccov
[1] 0.9214
$percslen
   97.5%
```

```
1.722824

$jcicov
[1] 0.9234
$jslen
[1] 1.772316
$adjcov
[1] 0.9118
$adjslen
[1] 1.772316
$asycov
[1] 0.9234
$asyslen
[1] 1.772316
```

```
source("http://parker.ad.siu.edu/Olive/sipack.txt")
args(tcisim)
function (n = 200, BB = 1000, nruns = 100, type = 1, dd = 3,
    eps = 0.1, shift = 9, alph = 0.05)
```

```
tcisim(n=200,type=1,nruns=5000)
#N(0,1) asymptotic scaled length 3.92
$tcicov
[1] 0.9536
$tslen
[1] 3.941903
```

Table 3.3. CI simulation, n=200, B=1000

e	t	shorth	perc	J	adj	asy
N,cov	0.9536	0.9548	0.95	0.9536	0.9538	0.9536
len	3.9419	3.9825	3.8898	3.9419	3.9419	3.9419
t,cov	0.9564	0.954	0.9462	0.9552	0.9314	0.9554
len	6.5853	6.6462	6.4992	6.5853	6.5853	6.5853
E,cov	0.9454	0.9466	0.9424	0.9466	0.9494	0.9464
len	3.9252	3.9598	3.8723	3.9252	3.9252	3.9252
U,cov	0.9482	0.9498	0.945	0.9484	0.9504	0.9484
len	2.2756	2.2974	2.2444	2.2756	2.2756	2.2756
MIX,cov	0.9538	0.9494	0.9372	0.9488	0.9122	0.9488
len	12.7838	12.9390	12.6396	12.7838	12.7838	12.7838
Pois,cov	0.9512	0.9514	0.9492	0.9516	0.9514	0.9516
len	11.8182	11.9360	11.6687	11.8182	11.8182	11.8182
Bin,cov	0.9492	0.9526	0.9498	0.9496	0.9496	0.9496
len	6.2278	6.2796	6.1424	6.2278	6.2278	6.2278
LN,cov	0.9302	0.9298	0.9284	0.9292	0.919	0.9292
len	8.2779	8.2839	8.1395	8.2779	8.2779	8.2779
Par,cov	0.9268	0.9292	0.9258	0.9288	0.9208	0.9286
len	1.7873	1.7892	1.7569	1.7873	1.7873	1.7873

```
$shorthcov
[1] 0.954
$shslen
[1] 3.98245
$perccov
[1] 0.95
$percslen
3.889827
$jcicov
[1] 0.9536
$jslen
[1] 3.941903
$adjcov
[1] 0.9538
$adjslen
[1] 3.941903
$asycov
[1] 0.9536
$asyslen
[1] 3.941903

tcisim(n=200,type=2,nruns=5000)
#t3 asymptotic scaled length 6.7896
$tcicov
[1] 0.9564
$tslen
[1] 6.585299
```

```
$shorthcov

[1] 0.954

$shslen

[1] 6.646193

$perccov

[1] 0.9462

$percslen

6.499239

$jcicov

[1] 0.9552

$jslen

[1] 6.585299

$adjcov

[1] 0.9314

$adjslen

[1] 6.585299

$asycov

[1] 0.9554

$asyslen

[1] 6.585299


tcisim(n=200,type=3,nruns=5000)

#Exp(1) asymptotic scaled length 3.92

$tcicov

[1] 0.9454

$tslen

[1] 3.925161
```

```
$shorthcov

[1] 0.9466

$shslen

[1] 3.959811

$perccov

[1] 0.9424

$percslen

3.872337

$jcicov

[1] 0.9466

$jslen

[1] 3.925161

$adjcov

[1] 0.9494

$adjslen

[1] 3.925161

$asycov

[1] 0.9464

$asyslen

[1] 3.925161


tcisim(n=200,type=4,nruns=5000)

#U(-1,1) asymptotic scaled length 2.2632

$tcicov

[1] 0.9482

$tslen

[1] 2.275636
```

```
$shorthcov

[1] 0.9498

$shslen

[1] 2.297353

$perccov

[1] 0.945

$percslen

2.244351

$jcicov

[1] 0.9484

$jslen

[1] 2.275636

$adjcov

[1] 0.9504

$adjslen

[1] 2.275636

$asycov

[1] 0.9484

$asyslen

[1] 2.275636


tcisim(n=200,type=5,nruns=5000)

#mixture asymptotic scaled length 12.9419

$tcicov

[1] 0.9538

$tslen

[1] 12.78379
```

```
$shorthcov
[1] 0.9494
$shslen
[1] 12.93904
$perccov
[1] 0.9372
$percslen
12.6396
$jcicov
[1] 0.9488
$jslen
[1] 12.78379
$adjcov
[1] 0.9122
$adjslen
[1] 12.78379
$asycov
[1] 0.9488
$asyslen
[1] 12.78379

tcisim(n=200,type=6,nruns=5000)
#Poisson(9) asymptotic scaled length 11.76
$tcicov
[1] 0.9512
$tslen
[1] 11.81819
```

```
$shorthcov
[1] 0.9514
$shslen
[1] 11.93603
$perccov
[1] 0.9492
$percslen
11.66868
$jcicov
[1] 0.9516
$jslen
[1] 11.81819
$adjcov
[1] 0.9514
$adjslen
[1] 11.81819
$asycov
[1] 0.9516
$asyslen
[1] 11.81819


tcisim(n=200,type=7,nruns=5000)
#bin(10,0.5) asymptotic scaled length 6.1981
$tcicov
[1] 0.9492
$tslen
```

```
[1] 6.227799

$shorthcov

[1] 0.9526

$shslen

[1] 6.279589

$perccov

[1] 0.9498

$percslen

[1] 6.14236

$jcicov

[1] 0.9496

$jslen

[1] 6.227799

$adjcov

[1] 0.9496

$adjslen

[1] 6.227799

$asycov

[1] 0.9496

$asyslen

[1] 6.227799


tcisim(n=200,type=8,nruns=5000)  #need n large

#lognormal(0,1) asymptotic scaled length 8.4719

$tcicov

[1] 0.9302

$tslen
```

```
[1] 8.277931

$shorthcov

[1] 0.9298

$shslen

[1] 8.283885

$perccov

[1] 0.9284

$percslen

8.139464

$jcicov

[1] 0.9292

$jslen

[1] 8.277931

$adjcov

[1] 0.919

$adjslen

[1] 8.277931

$asycov

[1] 0.9292

$asyslen

[1] 8.277931


tcisim(n=200,type=9,nruns=5000)  #need n large

#Pareto(1,0.25) asymptotic scaled length 1.8479

$tcicov

[1] 0.9268

$tslen
```

```
[1] 1.787255

$shorthcov

[1] 0.9292

$shslen

[1] 1.789192

$perccov

[1] 0.9253

$percslen

   97.5%

1.756939

$jcicov

[1] 0.9288

$jslen

[1] 1.787255

$adjcov

[1] 0.9208

$adjslen

[1] 1.787255

$asycov

[1] 0.9286

$asyslen

[1] 1.787255
```

```
source("http://parker.ad.siu.edu/Olive/sipack.txt")

args(tcisim)

function (n = 1000, BB = 1000, nruns = 100, type = 1, dd = 3,
```

Table 3.4. CI simulation, n=1000, B=1000

e	t	shorth	perc	J	adj	asy
N,cov	0.9518	0.9572	0.9498	0.9518	0.9522	0.9518
len	3.9247	3.9922	3.8996	3.9248	3.94248	3.9248
t,cov	0.9506	0.9542	0.9458	0.95	0.9382	0.9502
len	6.6915	6.8027	6.6477	6.6915	6.6915	6.6915
E,cov	0.9486	0.9512	0.9474	0.9484	0.9498	0.9486
len	3.9197	3.9838	3.8933	3.9197	3.9197	3.9197
U,cov	0.951	0.9516	0.948	0.951	0.9512	0.951
len	2.2664	2.23063	2.2531	2.2664	2.2664	2.2664
MIX,cov	0.9482	0.9506	0.9446	0.9472	0.9402	0.9472
len	12.8888	13.1219	12.8197	12.8888	12.8888	12.8888
Pois,cov	0.9492	0.954	0.951	0.9494	0.95	0.9492
len	11.7726	11.9810	11.7032	11.7726	11.7726	11.7726
Bin,cov	0.9452	0.947	0.9434	0.9452	0.9452	0.9452
len	6.2058	6.3115	6.1692	6.2058	6.2058	6.2058
LN,cov	0.9364	0.9386	0.935	0.938	0.9345	0.9378
len	8.4005	8.5239	8.3416	8.4005	8.4005	8.4005
Par,cov	0.948	0.951	0.946	0.9488	0.9432	0.949
len	1.8275	1.8542	1.8143	1.8275	1.8275	1.8275

```
    eps = 0.1, shift = 9, alph = 0.05)


tcisim(n=1000,type=1,nruns=5000)
#N(0,1) asymptotic scaled length 3.92
$tcicov
[1] 0.9518
$tslen
[1] 3.924763
$shorthcov
[1] 0.9572
$shslen
[1] 3.922175
$perccov
[1] 0.9498
$percslen
3.899624
$jcicov
[1] 0.9518
$jslen
[1] 3.924763
$adjcov
[1] 0.9522
$adjslen
[1] 3.924763
$asycov
[1] 0.9518
```

```
$asyslen
[1] 3.924763


tcisim(n=1000,type=2,nruns=5000)
#t3 asymptotic scaled length 6.7896
$tcicov
[1] 0.9506
$tslen
[1] 6.691495
$shorthcov
[1] 0.9542
$shslen
[1] 6.802742
$perccov
[1] 0.9458
$percslen
6.647684
$jcicov
[1] 0.95
$jslen
[1] 6.691495
$adjcov
[1] 0.938
$adjslen
[1] 6.691495
$asycov
[1] 0.9502
```

```
$asyslen
[1] 6.691495


tcisim(n=1000,type=3,nruns=5000)
#Exp(1) asymptotic scaled length 3.92
$tcicov
[1] 0.9486
$tslen
[1] 3.919668
$shorthcov
[1] 0.9512
$shslen
[1] 3.983847
$perccov
[1] 0.9474
$percslen
3.893267
$jcicov
[1] 0.9484
$jslen
[1] 3.919668
$adjcov
[1] 0.9498
$adjslen
[1] 3.919668
$asycov
[1] 0.9486
```

```
$asyslen
[1] 3.919668

tcisim(n=1000,type=4,nruns=5000)
#U(-1,1) asymptotic scaled length 2.2632
$tcicov
[1] 0.951
$tslen
[1] 2.266397
$shorthcov
[1] 0.9516
$shslen
[1] 2.306349
$perccov
[1] 0.948
$percslen
2.253059
$jcicov
[1] 0.951
$jslen
[1] 2.266397
$adjcov
[1] 0.9512
$adjslen
[1] 2.266397
$asycov
[1] 0.951
```

```
$asyslen
[1] 2.266397


tcisim(n=1000,type=5,nruns=5000)
#mixture asymptotic scaled length 12.9419
$tcicov
[1] 0.948
$tslen
[1] 12.8888
$shorthcov
[1] 0.9506
$shslen
[1] 13.12189
$perccov
[1] 0.9446
$percslen
12.81967
$jcicov
[1] 0.9472
$jslen
[1] 12.8888
$adjcov
[1] 0.9402
$adjslen
[1] 12.8888
$asycov
[1] 0.9472
```

```
$asyslen
[1] 12.8888


tcisim(n=1000,type=6,nruns=5000)
#Poisson(9) asymptotic scaled length 11.76
$tcicov
[1] 0.9492
$tslen
[1] 11.77258
$shorthcov
[1] 0.954
$shslen
[1] 11.98101
$perccov
[1] 0.951
$percslen
11.70317
$jcicov
[1] 0.9494
$jslen
[1] 11.77258
$adjcov
[1] 0.95
$adjslen
[1] 11.77258
$asycov
[1] 0.9492
```

```
$asyslen
[1] 11.77258



tcisim(n=1000,type=7,nruns=5000)
#bin(10,0.5) asymptotic scaled length 6.1981
$tcicov
[1] 0.9452
$tslen
[1] 6.205824
$shorthcov
[1] 0.947
$shslen
[1] 6.311533
$perccov
[1] 0.9434
$percslen
[1] 6.19205
$jcicov
[1] 0.9452
$jslen
[1] 6.205824
$adjcov
[1] 0.9452
$adjslen
[1] 6.205824
$asycov
```

```
[1] 0.9452

$asyslen

[1] 6.205824


tcisim(n=1000,type=8,nruns=5000)  #need n large

#lognormal(0,1) asymptotic scaled length 8.4719

$tcicov

[1] 0.9364

$tslen

[1] 8.400505

$shorthcov

[1] 0.9386

$shslen

[1] 8.523852

$perccov

[1] 0.935

$percslen

8.341591

$jcicov

[1] 0.938

$jslen

[1] 8.400505

$adjcov

[1] 0.9346

$adjslen

[1] 8.400505

$asycov
```

```
[1] 0.9378

$asyslen

[1] 8.400505


tcisim(n=1000,type=9,nruns=5000)  #need n large

#Pareto(1,0.25) asymptotic scaled length 1.8479

$tcicov

[1] 0.948

$tslen

[1] 1.827499

$shorthcov

[1] 0.951

$shslen

[1] 1.85423

$perccov

[1] 0.946

$percslen

   97.5%

1.814257

$jcicov

[1] 0.9488

$jslen

[1] 1.827499

$adjcov

[1] 0.9432

$adjslen

[1] 1.827499
```

```
$asycov
[1] 0.949
$asyslen
[1] 1.827499
```

CHAPTER 4

CONCLUSIONS

The simulations were done in *R*. See R Core Team (2016). Programs are in the collection of functions *sipack.txt*. See (http://parker.ad.siu.edu/Olive/sipack.txt). Table 1-4 were made with `tcisim` using the iid error distributions 1: N(0,1), 2: t_3, 3: exponential(1), 4: uniform(-1,1), 5: 0.1 N(0,1) + 0.9 N(0,100), 6: Poisson(9), 7: binomial(10,0.5), 8: lognormal(0,1), and 9: for Pareto(1,1/4) reporting the cov and len for the following t-type CI's; t CI (1.3), shorth CI (2.5), percentile CI (2.4), Johnson CI (2.3), skewness adjusted t CI (2.1), and asymptotic percentile t CI (2.2). The cov, or coverage, is the proportion of times the CI contained the mean μ. The average scaled CI lengths, or len, is the CI length multiplied by $\sqrt{n}$.

After running the 4 CI simulations, with sample sizes n=30, n=100, n=200, and n=1,000, the following can be observed. Recall, we are wanting a coverage or cov ≥ 0.94 with a small len or length. The distributions had better coverage as the sample size n got larger with the 2 highly skewed distributions, lognormal and Pareto, producing very poor results throughout. The coverage for the lognormal distribution had no cov ≥ 0.94 for all CI simulations and the Pareto distribution only saw cov ≥ 0.94 for the large sample size of n=1,000. This concurs with Hesterburg's findings. As for the other distributions, the tables show that t CI and Johnson CI were consistent in having a cov ≥ 0.94 within all of the sample sizes used. Not including the LN and Par distributions, of the 4 CI simulations, the 7 remaining distribution types produced CI's with coverage < 0.94 28 times. Of these 28 times, the t CI had 2, the shorth CI had 6, the percentile CI had 9, the Johnson CI had 1, the adjusted CI had 8, and the asymptotic CI had 2. The average lengths correlated with the average coverage. As the average coverage decreased, so did the average length. The most accurate Confidence Intervals were the t and Johnson with the Johnson CI being slightly better by < 1 %. It should be noted that the percentile CI improved the most as n increased. At n=1,000, though the t CI and Johnson CI presented consistent results throughout, the percentile CI had the shortest average length with coverage ≥ 0.94.

REFERENCES

[1] Chen, L. (1995), "Testing the Mean of Skewed Distributions," *Journal of the American Statistical Association* 90, 767-772

[2] Frey, J. (2013), "Data-Driven Nonparametric Prediction Intervals," *Journal of Statistical Planning and Inference,* 143, 1039-1048.

[3] Golam Kibria, B.M. (2006), "Modified Confidence Intervals for the Mean of the Asymmetric Distribution," *Pakistan Journal of Statistics,* 22, 109-120.

[4] Hall, P. (1988), "Theoretical Comparisons of Bootstrap Confidence Intervals," (with discussion), *The Annals of Statistics*, 16, 927-985.

[5] Hesterberg, T., (2014), "What Teachers Should Know About the Bootstrap: Resampling in the Undergraduate Statistics Curriculum," available from (http://arxiv.org/pdf//1411.5279v1.pdf). (An abbreviated version was published (2015), *The American Statistician*, 69, 371-386.)

[6] Johnson, N.J. (1978) "Modified t Tests and Confidence Intervals for Asymmetrical Populations," *Journal of the American Statistical Association*, 73, 536-544.

[7] Kleijnen, J.P.C., Kloppenburg, G., and Meeuwsen, F. (1986), "Testing the Mean of an Asymmetric Population: Johnson's Modified t Test Revisited," *Communications in Statistics: Simulation and Computation*, 15, 715-732.

[8] Moore, D.S. (2013), *The Basic Practice of Statistics,* 6th ed., W.H. Freeman, New York, NY.

[9] Olive, D.J. (2014), *Statistical Theory and Inference*, Springer, New York, NY.

[10] Pelawa Watagoda, L. C. R., and Olive, D.J. (2021), "Bootstrapping Multiple Linear Regression After Variable Selection," *Statistical Papers*, 62, 681-700. See (http://parker.ad.//siu.edu/Olive/ppboottest.pdf).

[11] R Core Team (2016), "R: a Language and Environment for Statistical Computing," R Foundation for Statistical Computing, Vienna, Austria, (www.R-project.org).

[12] Sutton, C.D. (1993), "Computer-Intensive Methods for Tests About the Mean of an Asymmetrical Distribution," *Journal of the American Statistical Association*, 88, 802-810.

VITA

Graduate School
Southern Illinois University

Daina McKinney

Mmlawnandlandscaping@gmail.com

Southern Illinois University Carbondale
Bachelor of Arts, Mathematics, May 2019

Research Paper Title:
Some t-Type Confidence Intervals

Major Professor: Dr. David J. Olive