Applied Mathematical Sciences, Vol. 6, 2012, no. 92, 4547 - 4560

A Comparison between Normal and Non-Normal Data in Bootstrap

Suliadi Firdaus Sufahani

Department of Science and Mathematics,
Faculty of Science, Technology and Human Development,
Universiti Tun Hussein Onn Malaysia, Parit Raja,
86400 Batu Pahat, Johor, Malaysia

Asmala Ahmad

Department of Industrial Computing,
Faculty of Information and Communication Technology,
Universiti Teknikal Malaysia Melaka,
Hang Tuah Jaya, 76100 Durian Tunggal, Melaka, Malaysia
asmala@utem.edu.my

Abstract

In the area of statistics, bootstrapping is a general modern approach to resampling methods. Bootstrapping is a way of estimating an estimator such as a variance when sampling from a certain distribution. The approximating distribution is based on the observed data. A set of observations is a population of independent and observed data identically distributed by resampling; the set is random with replacement equal in size to that of the observed data. The study starts with an introduction to bootstrap and its procedure and resampling. In this study, we look at the basic usage of bootstrap in statistics by employing R. The study discusses the bootstrap mean and median. Then there will follow a discussion of the comparison between normal and non-normal data in bootstrap. The study ends with a discussion and presents the advantages and disadvantages of bootstraps.

Keywords: Bootstrap, Resampling, Monte Carlo

1 Introduction

Bootstrap is well known as a resampling procedure. Resampling starts with drawing a sample from the population. For example, x = (x1, x2,…xn) is a sample. Then from the sample another sample, X*= (X1*, X2*,…Xn*), is drawn randomly with replacement. This technique is called resampling. Resampling, in other words, draws repeated samples from the given data. Bootstrapping and permutation are mostly used in statistics. Resampling is now widely used for confidence limits, hypothesis tests and other inferential problems. Resampling lets us analyse the sorts of data, even data with complex structures, for example regression models. However, which residuals should be resampled? This is one of the most confusing questions that arise in terms of regression prospects. Therefore, raw residuals leave one option and another one is studentized residuals, which are mostly involved in linear regressions. By using the studentized residual, it will make it easy for us to compare and run the results and it will not give us a major difference in actual practice.

2 Bootstrap Methods

Let us consider a certain situation with a common data, where a random sample $x = (x_1, x_2, \ldots x_n)$ from an unknown distribution F has been observed. Then we try to estimate the parameter of interest, $\hat{\theta}$ = t(F), which is on the basis of x. Then we calculate the estimate, $\hat{\theta}$ = t(x), for x. Therefore, in order to know how accurate $\hat{\theta}$ is we may need to use bootstrap. The bootstrap method is a computer based nonparametric technique for assigning measures of accuracy to sample estimates [2]. The technique allows the estimation of the sample distribution by using any simple method. Therefore, the main goal of bootstrap is to make an inference about the population parameter based on the sample statistic. Bootstrap belongs to frequentist statistics and not to Bayesian statistics. Bootstrapping is mostly used for estimating variance when sampling from an empirical distribution of the observed data. Mostly, the observations are from an independent and identically distributed population and can be implemented by constructing random sampling with replacement on the observed data set equal in size to the observed data set. Figure 1 gives an example of a bootstrap. “B” bootstrap samples are generated from the original data set. Each bootstrap sample has n elements, generated by sampling with replacement n times from the original data set. Bootstrap replicates, $s(x^{*1}), s(x^{*2}), \ldots s(x^{*B})$, are obtained by calculating the value of the statistic s(x) on each bootstrap sample. Finally, the standard deviation of the values $s(x^{*1}), s(x^{*2}), \ldots s(x^{*B})$ is our estimate of the standard error of s(x) [1]. The resampling can be carried out on a computer using a random number generator; in this case we use R [5]. Bootstrap is, therefore, a Monte Carlo technique, which is a numerical technique as opposed to an analytic technique, such as the *t*-test and F-test.

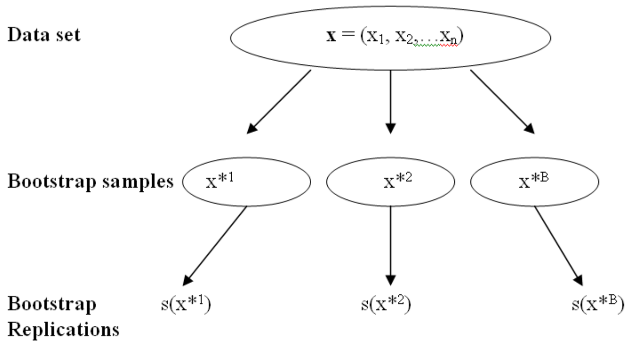

Fig. 1. Schematic of the bootstrap process for estimating the standard error of a statistic [1].

3 Comparison of Resampling Methods

Table 1 shows a comparison of the methods for testing the equality means of two populations [4].

Table 1: Comparison of resampling methods.

Permutation	**Rank (Wilcoxon)**	**Nonparametric Bootstrap**	**Parametric (*t*-test)**
Choose test statistic	Choose test statistic	Choose test statistic	Choose the test statistic whose distribution can be derived analytically
(e.g., sum of observations in first sample)	(e.g., sum of rank in first sample)	(e.g., difference between means of samples)	(e.g., Student's *t*)
Calculate statistic	Convert to ranks Calculate statistic	Calculate statistic	Calculate statistic
Are observations exchangeable?	Are observations exchangeable?	Are observations independent? With identical parameters of interest?	Are observations independent? Do they follow a specified distribution?
Derive permutation distribution from the combined sample	Use table of permutation distribution of rank	Derive bootstrap distribution: resample separately from each sample	Use tabulated distribution
Compare statistic with percentiles of distribution	Compare statistic with percentiles of distribution	Compare statistic with percentiles of distribution	Compare statistic with percentiles of distribution

Bootstrap is a relatively recent introduction, primarily because bootstrap is computationally intensive. Bootstrap, like the permutation test, requires a minimal number of assumptions and derives its critical values from the data at hand 0.

4 Mean Bootstrap

There are several methods for a single parameter and in our case we are considering the mean bootstrap. The methods are the percentile method, Lunneborg's method, traditional confidence limits and bootstrapped *t* intervals. The mean bootstrap is the simplest out of the bootstrapping procedures and the result will appear straightforward. In fact, it produces very nice confidence intervals on the mean and it is more sensible if we understand the mean in terms of statistics. Based on [1] along with a number of other researchers, he has come up with better limits. We develop an example for this case in order to explain this better. Firstly, we construct a set of data consisting of 50 random normal data and apply it to a simple linear regression [3]. Then we calculate the confidence interval for the data. The result is shown in Table 2.

Table 2: (a) Results from the example and (b) 90 % confidence interval for the mean statistic.

(a)

Mean Original Sample	Mean Bootstrap Sample	Bootstrap Bias	Bootstrap Standard Error
50.5	50.1766	-0.3234	2.802816

(b)

Lower Bound	Upper Bound
45.3755	54.704

Figure 2 shows the results of generating 95% confidence limits on the mean. The lower limit is 45.3755 units below the mean, while the upper is 54.704 units above the mean. We can also see that the distribution of the means is approximately normal, and the standard error of this bootstrapped distribution with its standard error is 2.802816.

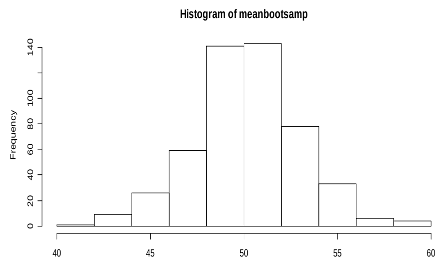

Fig. 2. Histogram for the example.

5 Median Bootstrap

The same thing goes for the median bootstrap; it uses the same method as the mean bootstrap, which is the percentile method, Lunneborg's method, traditional confidence limits or bootstrapped *t* intervals. We will not repeat this here; the result should be as straightforward as that for the mean bootstrap. We develop our own example for the case in order to better explain how to involve the median in bootstrapping. Firstly, we generate random data (for example, normal) and we use it as our raw data for the simple bootstrap and in this case we use 10 normal random data entries as our original data. Then we find the mean and the standard deviation of the original data. From the original data, we generate bootstrap samples by taking a certain amount (such as 20 in this example) of observations with 100 replications for each bootstrap sample with the mean and standard deviation. The mean of the data is considered to be the normal mean because it comes from a normal distribution of the data set. Then we calculate the median and standard error for each bootstrap and compare the original data with the resampling. As mentioned, we create a set of data consisting of 10 normal random data entries as shown in the table below. With a further calculation, we obtain a normal mean of 4.6 and the standard deviation is 2.875181 (Table 3).

Table 3: (a) Result from the set of normal random data and (b) mean and standard deviation.

(a)

Random Data	1	2	3	4	5	6	7	8	9	10

(b)

Mean	Standard Deviation
4.6	2.875181

The studies continue by constructing the bootstrap method in the data set. In order to obtain 20 bootstrap samples, the procedure should be based on the above data set, mean and standard deviation. By taking 100 replications for each bootstrap sample, we set up the mean above as a normal mean and the standard deviation as a normal standard error in order to construct the bootstrap method. The table below shows the first bootstrap sample of the 20 samples with 100 replications. These methods are called resampling.

Table 4: 100 replications for the first bootstrap sample out of 20 bootstrap samples.

5	6	4	2	6	6	4	4	5	4
-1	6	5	1	6	3	0	3	7	6
0	5	2	7	2	8	5	5	6	7
6	0	10	1	6	0	4	8	5	4
4	7	7	3	4	5	6	3	2	0
4	4	7	4	6	2	5	5	0	7
8	1	5	5	7	3	6	4	2	5
7	5	11	6	6	4	5	3	3	4
7	5	2	1	8	6	5	6	4	4
8	0	4	8	6	1	6	1	6	5

Then we calculate the median for each bootstrap sample as shown in Table 5. Once we obtain the median for each bootstrap sample, we continue with the next procedure, which is calculating the standard deviation of the bootstrap distribution. Based on the resample median, we can find the standard deviation.

Table 5: (a) Median for all the 20 bootstrap samples that we generate and (b) the standard deviation.

(a)

1	2	3	4	5	6	7	8	9	10
4	4.5	4	5	4	5	5	5	5	4
11	12	13	14	15	16	17	18	19	20
5	5	5	5	5	4	5	5	6	5

(b)

Standard Deviation	0.2236068

In order to see the bootstrap median, we construct a histogram for the distribution of the medians (Figure 3).

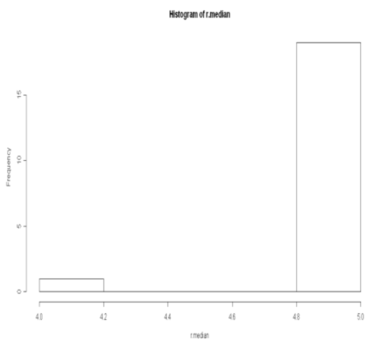

Fig. 3. Distribution bootstrap median based on the study.

From the histogram, we can see that most of the median lies on the value of 5 rather than the value of 4. Furthermore, the standard deviation of the bootstrap sample is 0.2236068. For further studies, we can put all these steps into a single function where all we would need to do is to specify the data set and how many times we want to resample in order to obtain the adjusted standard error of the median.

6 Monte Carlo Looping With Bootstrap In A Simple Linear Case For Normal And Non-Normal Data

An exploratory analysis of the summary statistics for the response variables and explanatory variables is carried out to calculate the simple regression case for the normal and non-normal data in bootstrap. To make the interpretation easy, a few graphical representations using a histogram are also carried out at the end of each result. Before going for statistical modelling, the correlations between response and explanatory variables are examined by illustrating a few scatter plots. The data given is already in a linear relationship between the response and explanatory variables. An exploratory analysis of the data is carried out using the R statistical tool. We are dealing with a simple linear regression. For a simple regression the statistical model can be expressed as follows:

$$y_i = \beta_0 + \beta_1 x_i + e_i, \quad e_i \sim N(0,\sigma^2_e) \tag{1}$$

where y_i is the response of interest, β_0 is the overall mean, x_i is the independent variable and e_i is the normally distributed residual term. For the estimation coverage of the bootstrap and classical methods, we perform the correlation model in the simple linear regression to obtain the property of beta-one (β_1), where β_1 is the slope. However, we are using Monte Carlo looping 100 times. The result will indicate the true value of the parameter in the classical method and bootstrap method. Then we calculate the estimation coverage of the confidence interval and the general formula is shown below:

$$\text{Estimation Coverage of CI} = \left(\frac{1}{M}\right)\sum I_m \tag{2}$$

M = number of times there is Monte Carlo looping

$\sum I_m$ = the sum of times when the fixed value of β_1 lies within the lower and upper bounds or, in other words, it contains a true parameter.The estimation coverage is carried out for the normal and non-normal data and from there we compare the differences between the two distributions, using the bootstrap and classical methods. In this example, firstly, we fix up the value of beta-one (β_1) with a certain value. Then we used 100 random data entries (for example, normal or non-normal, such as gamma) with a mean of 100 and a standard deviation of 50 as our original data for independent variables (x) and another 100 normal random data entries with a normal mean of 0 and a standard deviation of 50 for dependent variables (y). In this example, y is the dependent variable and x is the independent variable. We set the calculation by generating 200 bootstrap samples where each sample is observed with 100 replications. From there we find the 95% confidence interval for the mean statistics for the classical method and the bootstrap method. Then we calculate the estimation coverage by checking whether the fixed value of β_1 lies within the lower and upper bounds for the bootstrap and classical methods. As mentioned earlier, we give the value of 1 when it lies and 0 otherwise

6.1 Normal Data

Firstly, we fixed the value of beta-one ($\boldsymbol{\beta_1}$) as 0.3 from the population. Therefore, the equation is as follows:

$$y_i = \beta_0 + 0.3x_i + e_i, \quad e_i \sim N(0,\sigma^2_e) \qquad (3)$$

Secondly, we set up a sample data set, which is the independent variable (x) with normal random data. The sample set consists of 100 data entries for x and we want to use the mean top to describe the centre of the data. We run Monte Carlo looping 200 times; so there are 200 sample sets with a different data set for every loop. The data in Table 6 is one of the sample sets out of 200 sets.

Table 6: **O**ne of the examples of the data set for x.

-0.278	0.058	0.987	-0.765	7.998	12.974	-5.998	-0.35	0.157	-0.833
1.011	-0.900	1.907	7.998	3.908	1.873	10.243	1.211	0.603	-0.912
1.196	1.067	0.759	56.986	12.908	-32.861	12.098	-0.659	-0.076	1.247
-1.009	-0.265	-0.876	12.765	87.095	-67.982	0.145	-0.573	0.711	-0.103
0.879	1.309	-1.765	0.784	-27.872	44.321	74.213	67.987	1.364	12.098
1.564	-1.112	2.875	65.543	-9.087	-0.673	98.028	17.884	89.442	54.008
-0.345	0.788	5.981	35.932	5.975	-99.023	59.763	-1.009	-35.098	19.078
-0.654	0.429	4.100	-16.936	3.865	56.985	47.328	16.429	-56.998	0.987
0.167	-0.797	-5.231	78.338	-6.981	-17.843	49.467	83.002	0.364	34.587
0.050	-1.20	0.756	-1.763	0.976	-63.209	0.764	-0.286	-0.891	-0.466

Thirdly, we set up a sample data set, which is the dependent variable (y) with normal random data. The sample set consists of 200 data entries for y and we want to use the mean top to describe the centre of the data. We run Monte Carlo looping 200 times; so there will be 200 samples set with a different data set with every loop. The data in Table 7 is for one of the sample sets out of 200 sets.

Table 7: One of the examples of the set of data for y.

73.586	105.603	66.304	114.935	20.785	196.067	215.079	157.364	13.353	213.675
201.366	37.3447	57.017	171.174	125.463	125.616	133.205	194.818	173.922	247.951
92.717	112.443	152.764	72.760	66.644	219.722	186.013	8.291	209.785	192.843
79.576	185.700	94.843	93.094	171.878	70.877	141.636	45.366	173.922	117.795
15.268	-14.105	135.989	218.542	70.877	70.877	72.959	284.018	53.288	142.644
15.268	91.355	111.646	129.773	4.120	195.769	45.535	-25.517	53.288	142.644
26.449	76.382	215.630	156.986	106.170	65.775	99.969	4.955	-19.497	127.430
-59.724	192.107	50.378	74.778	106.170	168.468	167.576	153.017	113.437	287.776
90.200	104.754	112.492	64.250	239.187	23.818	131.103	109.365	-34.964	-18.050
-49.795	120.135	-10.081	-53.994	321.957	214.047	201.541	117.073	247.951	174.954

From the data above, as mentioned earlier, we determine that x is an independent variable and y is a dependent variable. The study continues by generating 200 bootstrap samples for both x and y and each bootstrap sample has 100 replications, with a significant level of 5%. In this section, we perform the correlation model in the simple linear regression and obtain the property of β_1, where β_1 is the slope. However, we are using Monte Carlo looping 200 times. The results are indicated using the classical method and bootstrap methods. The results below show an example of the lower and upper bounds of β_1 for the first set out of the 200 sets. Table 8 shows that, by using the classical method, we set up a calculation to indicate whether the fixed value of β_1 lies between the lower and upper bounds for the classical method and bootstrap methods.

Table 8: One of the examples of the lower and upper bounds for the 95% confidence interval using (a) the classical method and (b) the bootstrap method.

(a)

	β_1
Lower Bound	-0.4854653
Upper Bound	0.7736806

(b)

	β_1
Lower Bound	-0.1497580
Upper Bound	0.7855656

We have run the Monte Carlo looping 200 times and from the simulation, in 198 out of 200 times, the fixed value of β_1 lies between the lower and upper bounds for the classical method and, in 187 out of 200 times, the fixed value of β_1 lies between the lower and upper bounds for the bootstrap method. |Therefore, from the information, the calculation is as follows:

For the classical method,

$$\text{Estimation Coverage of CI} = \left(\frac{1}{M}\right)\sum I_m = \left(\frac{1}{200}\right)*198 = 0.96 \qquad (4)$$

For the bootstrap method,

$$\text{Estimation Coverage of CI} = \left(\frac{1}{M}\right)\sum I_m = \left(\frac{1}{200}\right)*187 = 0.935 \qquad (5)$$

From the results, if we fix the value of the β_1, which is 0.3, we can see that the classical method covers 96% of the fixed value within it lower and upper bounds and the bootstrap method covers 93.5% of the fixed value within its lower and upper bounds. It seems that, for normal cases, both the classical method and the bootstrap method have a value for the estimation coverage that is nearly the same as that for the 95% confidence interval. This is true if we base this on the rules below (for example, the Poisson distribution):

$\alpha = 0.05$ $\qquad n = 200$ $\qquad n\alpha = \lambda = 10$ $\qquad x \approx \text{Poisson}(\lambda)10$

$$P(\text{Poisson}(\lambda) \leq 10) = \sum e^{-\lambda}\left(\frac{\lambda^r}{r!}\right) \sum e^{-\lambda}\left(\frac{\lambda^r}{r!}\right)$$ and

$$P(\text{Poisson}(\lambda) \geq 10) = 1 - \sum e^{-\lambda}\left(\frac{\lambda^r}{r!}\right)$$

Based on the rule, the value of the estimation coverage should be close to the value of the percentage of the confidence interval. Then a further study is carried out by changing the value of the fixed β_1 (Table 9).

Table 9: Future test of the estimation coverage by changing β_1 value.

	Estimation Coverage Classical Method	Estimation Coverage Bootstrap Method
Normal (β_1 = 0.1)	0.96	0.935
Normal (β_1 = 0.15)	0.98	0.945
Normal (β_1 = 0.2)	0.96	0.93
Normal (β_1 = 0.25)	0.955	0.92
Normal (β_1 = 0.35)	0.965	0.9

The table above shows that, even though we change the value of β_1, the answer for both methods is still close to the percentage of the confidence interval. It proves that the studies are correct and reliable.

6.2 Non-Normal Data (Gamma-Distribution)

Firstly, we fix the value of β_1 at 0.3 for the population. Therefore, the equation is as follows:

$$y_i = \beta_0 + 0.3x_i + e_i \qquad X \sim \text{Gamma}(k,\theta) \qquad (6)$$

As we know, the pdf for the gamma distribution is:

$$f = \left(\frac{1}{\left(S^a \Gamma(a)\right)} \right) x^{a-1} e^{\frac{-x}{S}} \qquad (7)$$

where a is the shape and s is the scale. Therefore, the mean and variance for the gamma distribution are E(X) = a*s and Var(X) = a*s^2. In this study we set the value of the mean as 10 and the value of variance as 20. Secondly, we set up sample data set with the independent variable (x) and non-normal random data. The sample set consists of 100 data entries for x and we want to use the mean top to describe the centre of the data. We run Monte Carlo looping 200 times, so there will be 200 sample sets with a data set with every loop. The data in Table 10 is for one of the sample sets out of 200 sets.

Table 10: One of the examples of the data sets for x.

11	4	6	15	21	23	16	9	18	5
2	21	4	5	19	5	11	8	17	7
12	14	24	1	3	14	25	25	16	15
22	9	7	11	12	20	13	3	10	8
18	3	13	11	16	19	7	23	8	1
4	5	24	10	1	6	10	6	14	1
23	23	24	22	4	7	12	22	10	21
18	18	9	1	22	3	1	8	18	19
4	14	21	9	11	9	25	17	1	20
23	6	3	9	19	13	16	19	18	17

Thirdly, we set up sample data set with the dependent variable (y) and non-normal random data. The sample set consists of 100 data entries for y and we want to use the mean top to describe the centre of the data. We run Monte Carlo looping 200 times, so there will be 200 sample sets with a different data set for every loop. The data in Table 11 is one of the sample sets out of 200 sets.

Table 11: One of the examples of the set of data for y.

19	6	20	19	2	6	5	15	17	21
16	24	4	16	15	5	1	5	4	12
4	6	13	18	18	20	21	6	20	3
9	11	23	15	20	9	1	19	3	15
16	14	24	6	17	20	17	9	3	1
8	11	18	17	22	12	4	25	9	5
11	23	24	25	5	1	19	9	17	5
2	20	15	13	4	12	5	19	20	12
4	10	23	22	3	4	4	15	3	14
9	15	4	20	2	7	17	20	10	17

From the data above, as mentioned earlier, we determine that x is an independent variable and y is a dependent variable. The study continues by generating 200 bootstrap samples for both x and y and each bootstrap sample has 100 replications, with a significance level of 5%. In this section, we perform the correlation model in a simple linear regression and obtain the property of β_1, where β_1 is the slope. However, we are using Monte Carlo looping 200 times. The results are indicated for the classical method and bootstrap methods. The results in Table 12 shows an example of the lower and upper bounds of β_1 for the first set out of 200 sets.

Table 12: One of the examples of the lower and upper bounds for the 95% confidence interval using (a) the classical method and (b) the bootstrap method.

(a)

	β_1
Lower Bound	-0.491317
Upper Bound	0.3697785

(b)

	β_1
Lower Bound	-0.4927856
Upper Bound	0.4019495

Based on the table above, we set up a calculation to indicate whether the fixed value of β_1 lies between the lower and upper bounds for the classical method and bootstrap method. We run the Monte Carlo looping 200 times and from the simulation, in 189 out of the 200 times, the fixed value of β_1 lies between the lower and upper bounds for the classical method and, in 181 out of 200 times, the fixed value of β_1 lies between the lower and upper bounds for the bootstrap method. Therefore, from the information, the calculation is as follows.

For the classical method,

$$\text{Estimation Coverage of CI} = \left(\frac{1}{M}\right)\sum I_m = \left(\frac{1}{200}\right)*189 = 0.945 \qquad (8)$$

For the bootstrap method,

$$\text{Estimation Coverage of CI} = \left(\frac{1}{M}\right)\sum I_m = \left(\frac{1}{200}\right)*181 = 0.905 \qquad (9)$$

From the results, if we fix the value of the $\beta_{1,}$ which is 0.3, we can see that the classical method covers 94.5% of the fixed value in it lower and upper bounds. However, the bootstrap method covers 90.5% of the fixed value in its lower and upper bounds. It seems that, for non-normal cases, the classical method has a high range of lower and upper bounds when compared to the bootstrap method. It seems that, for normal cases, both the classical method and bootstrap method have a value for the estimation coverage that is nearly the same as that for 95% of the confidence interval even though we are dealing with non-normal data, which is the gamma distribution. This is true if we base this on the rules below, which is the same as for our discussion of the earlier results (for example, the Poisson distribution), which is the same as we discussed earlier in the normal data section. Based on the rule, the value of the estimation coverage should be close to the value of the percentage of the confidence interval. Then we carry out a further study by changing the value of the scale (s) and shape (a) in order to see the estimation coverage for each method (Table 13).

Table 13: Future test of the estimation coverage by changing the value of the shape (a) and scale (s).

Value of (a)	Value of (s)	Mean a*s	Variance $a*s^2$	Estimation Coverage Classical Method	Estimation Coverage Bootstrap Method
1	5	5	25	0.95	0.895
2	4	8	32	0.92	0.94
3	3	9	27	0.95	0.915
4	2	8	16	0.96	0.895
5	1	5	5	0.92	0.905

The table above shows that, even though we change the value of the shape (a) and scale (s), the answer for each method is still close to the percentage of the confidence interval. It proves that the studies are correct and reliable.

7 Conclusion and Discussion

Based on the study, a general explanation has been given regarding the background to bootstrap. All the procedures of bootstrap have been discussed to provide further understanding of the application of bootstrap as well as a simple calculation for bootstrapping using R. An analysis of the data has been conducted in order to find the mean and median for bootstrap. From the results we can see that bootstrap gives a better result than the original value if we compare both results. Furthermore, the authors are using the Monte Carlo technique and looping. From there we say can that, whether we are using normal or non-normal data, both the estimation coverage for bootstrap and the estimation coverage for the classical method are close to the value of the confidence interval. Furthermore, if we use non-normal data, at the end of bootstrapping we can see that the distribution appears to be nearly normal. In the future, we should consider using bootstrapping in the Monte Carlo technique and looping in the Bayesian diagnostic or time series. From there we can see the pattern of the distribution and the implementation of the results. The study should be conducted on both normal and non-normal data. The advantage of bootstrapping is that the result is simple and straightforward even when it involves a complex parameter. However, the disadvantage is that it does not provide a sample guarantee and has a tendency to make certain important assumptions, such as the independence of a sample.

References

[1] B. Efron, R. Tibshirani, An Introduction to the Bootstrap, Chapman & Hall/CRC, Florida, 1993.

[2] G. David, Bootstrapping and the Problem of Testing Quantitative Theoretical Hypothesis, Middle East Technical University, 2002.

[3] J. Fox, Bootstrapping Regression Model, Brown University, 2002.

[4] P. Good, Permutation Test, 2nd ed, Springer, New York, 2000.

[5] P. Good, Resampling Method, 3rd ed, Birkhauser, Boston, 2006.

Received: April, 2012