International Journal of Information Technology and Business Management
29th October 2014. Vol.30 No.1

ISSN 2304-0777 www.jitbm.com

ON FINDING THE SMALLEST REPLICATION NUMBER OF BOOTSTRAP CONFIDENCE INTERVAL USING COMPUTATIONAL SIMULATION

Zaturrawiah Ali 0mar [1], Noraini Abdullah [2], Zainodin Jubok[2] and Tan Wei Hsiang[3]
[1] Mathematics with Computer Graphics Programme, [2] Mathematics with Economics Programme , [3]Industrial Chemistry Programme.
[1,2,3]Mathematics and Statistical Applications Research Group, Faculty of Science and Natural Resources, Universiti Malaysia Sabah

Email: zatur@.ums.edu.my

Abstract

A preliminary experiment was conducted to investigate if a smaller number of bootstrap replication B for confidence interval was able to be identified through computational simulation. Interest was particularly on small data sets and in this simulation, a data set of Cadmium (Cd) concentration was used with 28 observations. A stopping criteria was identified and standard errors when B = 1000 and B = 2000 were used as benchmark. The results show that the stopping criteria were satisfied on specified cases of comparison up to d decimal point. As d increased, so did the number of B. For the CD case specifically, we found that the range of 100 to 1600 were adequate.

Keywords: *Small sample size, bootstrap replications, confidence interval, cadmium.*

INTRODUCTION

This study looked into the question of how many bootstrap replication – *B* for the statistic of interest $\hat{\theta}^*$ was required for conducting 90 – 95 percent confidence intervals. Bootstrap is a computer intensive method that does resampling with replacement and statistical conclusion is drawn from the bootstrap sampling (Henderson, 2005). As stated by Carpenter and Bithell (2000) that most practitioners suggest *B* should be between 1000 and 2000. This is to ensure inadequate bootstrap sampling as explained by Efron (1987). Yet, getting a *B* as small as possible is still preferable as replication will not add any more information to the existing data set. Replication however, can reduce the standard error of the mean (Efron, 1987, Wang *et al*., 2013). Efron & Tibshirani (1986) have demonstrated this by the coefficient of variation (CV) of estimating the true standard error and further explain that the bootstrap estimation of standard error will have a larger CV due to random sampling.

The question on how many bootstrap replications is not new. There has been few other researches, looking at the same question and came to their conclusion. Efron & Tibshirani (1986) for example, recommends *B* = 1000 as the rough minimum and further states that *B* should be in the order of 1000 (Efron, 1987). Fisher & Hall (1991) on the other hand has come with an exact calculation of finding *B* for small sample size of $n < 7$ and shows that when $n = 7$, $B = 1716$. Other similar works on different cases can also be seen in Blair (1992), Pattengale *et al*. (2010) and Wang *et al*. (2013).

Our approach to answer the question was by fully utilizing the power of computers; simulate the bootstrap resampling process repeatedly, a large number of times. The data used was on Cadmium (Cd) concentration found in tree barks of the

Cinnamomum Inners along the main entrance road of Universiti Malaysia Sabah. The outer barks of the trees were peeled off and the inner were taken as samples. It was then dried, grinded and digested using acid digestion (Skoog *et al.*, 2007) then analysed using Inductively Coupled Plasma Optical Emission Spectra (ICP-OES) to obtain its concentration.

Cd is one of the metals that can be found naturally in soil. Most uncontaminated soils are expected to contain <1 mg/kg of Cd (Page & Bingham, 1973). Cd can be toxic if it exceeds its maximum permissible limits (MPLs). According to Krejpcio *et al.* (2007), the MPLs of Cd in spices and herbs is 0.1 mg/kg. Meanwhile, according to Ross, (1994) and Kabata-Pendias & Pendias (2001), the normal limits of Cd concentration in plants are between 0.2 mg/kg to 0.8 mg/kg. It is considered toxic when it is between 5 mg/kg to 30 mg/kg.

METHOD

The investigation required a stopping criteria for identifying the smallest *B* and here we used the formula looking at bootstrap standard deviation $\sigma^* = \frac{S}{\sqrt{n}}$ where *S* is the sample standard deviation and *n* is the sample size (Efron & Tibshirani, 1986). Since the true σ is not known, then *S* was used as an estimate. To ensure that *S* was a good estimator, normality and randomness tests were conducted using `Shapiro.test()` (Shapiro-Wilk normality test) and `runs.test()` functions in R version 3.0.1 (R Core Team, 2013).

The comparison on σ^* was on its rounded value in *d* number of digits where the rounding function used was `round(valueToRound, d)` (R Core Team, 2013). Here *d* was set to be in cases of 2, 3, 4, 5 and 6. The range of *B* was from 30 to 5000. The `boot()` function then would be iterated with the interval of 1, starting from B = 30 until the stopping criteria was satisfied or, until *B* = 5000. Only when the stopping criteria was satisfied, the *B* value of that iteration (*B**) was saved. This was done repeatedly until 10 000 times for all *d* cases. Since our interest was to find a smallest number, the mean and median of *B** will be considered and these statistics will produced a single value (no ties, unlike mode). If the values were in fraction, it will be rounded up. Another bootstrap will be conducted based on the values. The observed $\bar{X}^*$, σ^* and $\sigma^*_{\bar{X}}$ then were saved.

To know how far (or near) was the value of the bootstrap estimated standard error with benchmark, the $\sigma^*_{\bar{X}}$ was compared with the $\sigma^*_{\bar{X}_B}$ when *B* = 1000 and *B* = 2000 using the equation $\sigma^*_{\bar{X}} - \sigma^*_{\bar{X}_B} / \sigma^*_{\bar{X}_B}$. The flow chart in Figure 1 summarises the simulation concentrating on the loops of the processes.

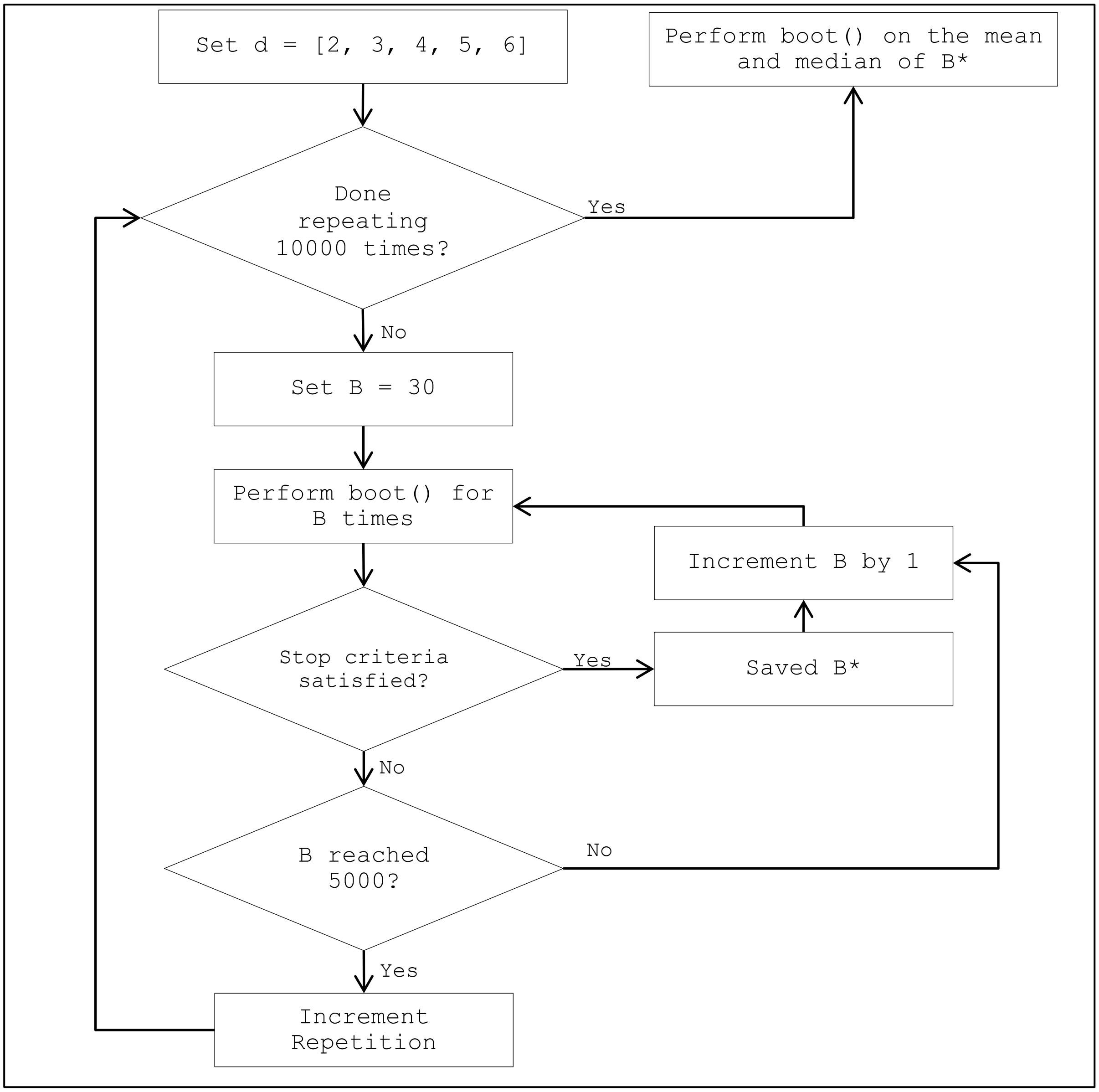

FIGURE 1. Flow Chart of Simulation Algorithm

RESULTS AND DISCUSSION

28 observations of Cd concentration were gathered with the following statistical information as shown in Table 1.

TABLE 1. Descriptive Statistic of Cd

N	Mean$(\bar{X})$	Standard Deviation (S)	Standard Error $(\sigma_{\bar{X}})$
28	0.25	0.12	0.02

Shapiro-Wilk normality test was conducted giving *p-value* = 0.573 ($> \alpha = 0.05$), thus accepting the null hypothesis that the sample follows a normal distribution. A run test was also conducted using function `runs.test()` from the `randtests` package in R. The results gave *p-value* = 0.4411 ($> \alpha = 0.05$), therefore accepting the null hypothesis that the sample data was random. Based on both tests, we concluded that the sample can be a good representative of the population. The histogram and the QQ plot of the data set can be referred in Figure 2.

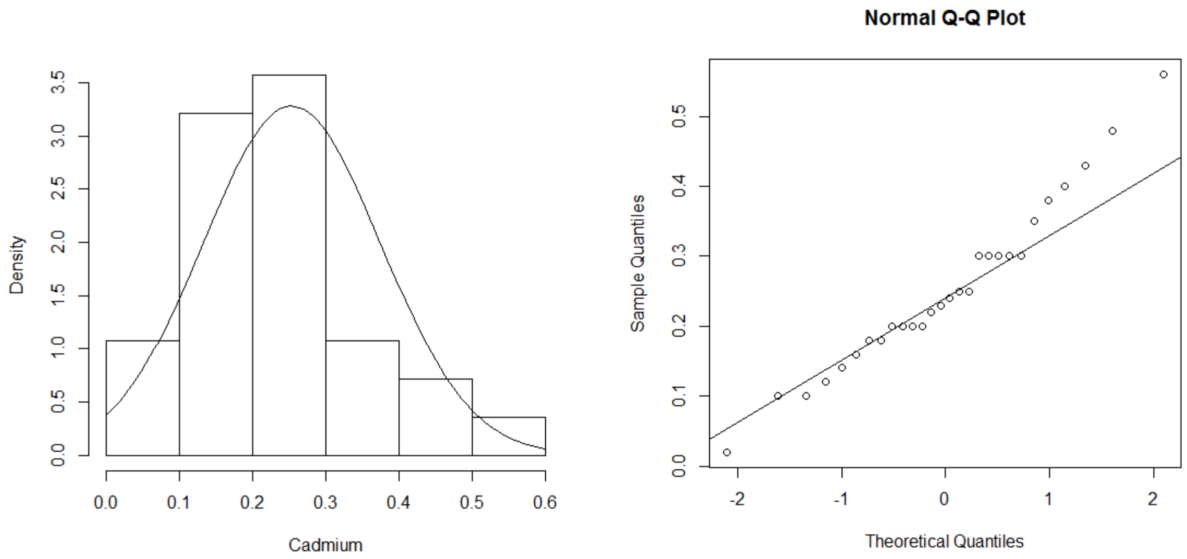

FIGURE 2. Histogram and QQ Plot of Cadmium Concentration with 28 observations

Results from conducting bootstrap when *B* = 1000 and *B* = 2000 then calculated (see Table 3) as a benchmark for comparison with the simulation results shown in Table 2.

TABLE 2. Results of Simulation

Rounded digits	*B*		Mean	Standard Deviation	Standard Error	$(\sigma^*_{\bar{X}} - \sigma^*_{\bar{X}_{B=1000}}) / \sigma^*_{\bar{X}_{B=1000}}$	$(\sigma^*_{\bar{X}} - \sigma^*_{\bar{X}_{B=2000}}) / \sigma^*_{\bar{X}_{B=2000}}$
d = 2	Mean	30	0.2554048	0.0246433	0.0044992	5.39	8.11
	Median	30	0.2554048	0.0246433	0.0044992	5.39	8.11
d = 3	Mean	36	0.2522024	0.0192158	0.0032026	3.55	5.49
	Median	34	0.2488550	0.0201217	0.0034508	3.90	5.99
d = 4	Mean	81	0.2481966	0.0229305	0.0025478	2.62	4.16
	Median	71	0.2522284	0.0218194	0.0025895	2.68	4.24
d = 5	Mean	327	0.2531684	0.0229904	0.0012714	0.80	1.57
	Median	276	0.2512164	0.0225958	0.0013601	0.93	1.75
d = 6	Mean	1581	0.2535610	0.0226246	0.0005690	-0.19	0.15
	Median	1313	0.2528784	0.0216928	0.0005987	-0.15	0.21

TABLE 3. Results when *B* = 1000 and *B* = 2000

	***B* = 1000**	***B* = 2000**
Mean $\bar{X}^*$	0.2541025	0.2529234
Standard Deviation σ^*	0.0222813	0.0220851
Standard Error $\sigma^*_{\bar{X}}$	0.0007046	0.0004938

The results analysis seems to indicate an exponential growth of the B values as the rounded digits increased (see Figure 3). This was true for both mean and median values, although mean values were slightly higher compared to the median.

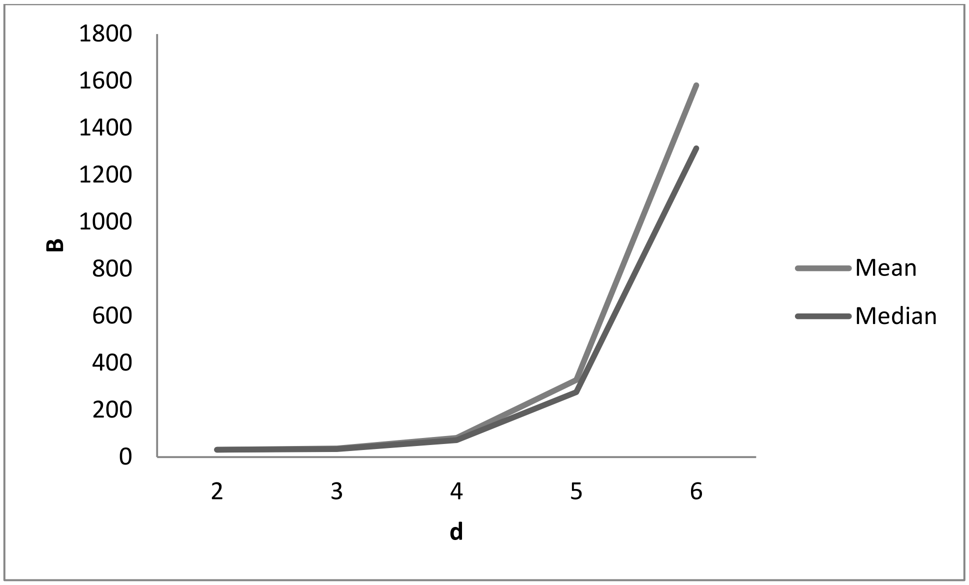

FIGURE 3. Mean and Median values of *B* over *d* number of rounded digits

As comparisons were started when the value of σ^* was rounded to 2 decimal points (2 digits), the mean and median values of *B** was 30. This somehow conforms to the minimum number of samples – *n*, where most textbooks suggest that *n* should be at least 30. The difference ratio however, was more than 5 times greater compared with *B* = 1000 and more than 8 times greater when *B* = 2000.

When *d* = 6, the mean and median of *B** exceeded 1000 but still did not exceed 2000 and its respective standard error ratios were 0.15 and 0.21 greater than the 2000 benchmark. Because of the exponential growth, this might not hold true when d > 6.

For the Cd case study, the smallest number of bootstrap replication will depend on the number of significant digits (accuracy) required by the researcher and we suggest that the smallest number should be to its nearest hundredth. We also believe that the mean values were more significant than the median. From the results, if *d* = 2 case was excluded, all of the mean values were greater when compared to the medians. This indicated that the *B** distribution was heavy towards its right. Therefore, for the Cd cases in particular, the significant suggested range of the smallest number of bootstrap replicates was between 100 – 1600 for 4 – 6 significant digits.

For constructing the confidence interval, Carpenter and Bithell (2000) has provided a guide on choosing a bootstrap confidence interval method and for the Cd study, the bootstrap method is the Bias Corrected and Accelerated method (BCa). BCa of 90% and 95%, then was conducted in 100, 1000, 1600 and 2000 replications (1000 and 2000 is the min and max values from the benchmark range;

100 and 1600 is the min and max values of the new suggested range for the case study) and the results are as in Table 4. From the BCa results, all interval estimates were below the dangerous MPLs.

TABLE 4. BCa results

B	90%		95%	
100	0.2221	0.2876	0.2029,	0.2957
1000	0.2189	0.2929	0.2143	0.2986
1600	0.2171	0.2929	0.2106	0.3010
2000	0.2179	0.2918	0.2104	0.2982

CONCLUSION

The investigation on finding the smallest *B* was conducted through computer simulations and the main algorithm of the simulation flow can be seen in Figure 1. The simulations required a stopping criteria and the standard exercise on finding the population standard deviation was used on a data set of Cd concentration. Therefore, it was essential that the data set (especially small data set) must first conform to be random and normally distributed. We also concluded that the mean values of *B** were more significant. Depending on the requirement of the researcher, we also found that *B* as small as 30 seems to be enough for accuracy up to 2 significant digits. The standard error of *B* = 30 however was 5 and 8 times greater as compared to the benchmarks (*B* = 1000 and *B* = 2000). For the Cd case, we suggest that *B* should be between 100 to 1600. A further calculation on finding the confidence interval using BCa method for the selected *B* values produces ranges that were within the normal limits.

REFERENCES

1. Blair Hedges, S. (1992). The Number of Replications Needed for Accurate Estimation of The Bootstrap p Value in Phylogenetic Studies. *Menopause (New York, N.Y.)*, *9*(**2**), 366–369.
2. Efron, B. (1987). Better Bootstrap Confidence Intervals. *Journal of the American Statistical Association*, *82*(397), 171–185.
3. Efron, &, B. & Tibshirani., R. (1986). Bootstrap Methods for Standard Errors, Confidence Intervals an Other Measures of Statistical Accuracy. *Statistical Science*. Vol **1**, No.1,54 – 77.
4. Carpenter, J., & Bithell, J. (2000). Bootstrap confidence intervals : when , which , what ? A practical guide for medical statisticians. *Statistics in Medicine*, (August 1999), 1141–1164.
5. Fisher, N. I., & Hall, P. (1991). Bootstrap algorithms for small samples. *Journal of Statistical Planning and Inference*, *27*(2), 157–169.
6. Henderson, A. R. (2005). The Bootstrap: A technique for data-driven statistics. Using computer-intensive analyses to explore experimental data. *Clinica Chimica Acta*, 359 1 – 26.
7. Kabata-Pendias, A. & Pendias, H. (2001). *Trace Elements in Soils & Plants*, 3rd ed. CRC Press, LLC, Boca Raton,Florida.
8. Krejpcio, Z., Król E., & Sionkowski, S. (2007). Evaluation of Heavy Metals Contents in Spices and Herbs Available on the Polish Market. *Polish Journal of Environmental Study*, **16**(1): 97-10.
9. Pattengale, N. D., Alipour, M., Bininda-Emonds, O. R. P., Moret, B. M. E., & Stamatakis, A. (2010). How many bootstrap replicates are necessary? *Journal of Computational Biology : A Journal of Computational Molecular Cell Biology*, *17*(3), 337–54.
10. R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/
11. Ross, S.M. (1994). Sources and forms of potentially toxic metals in soil-plant systems. In: S.M. Ross, (Eds.) *Toxic Metals in Soil-Plant Systems*, Chicester: John Wiley. 3-25.
12. Skoog, D. A., Holler, F. J., & Crouch, S. R. 2007. Principles of Instrumental analysis. 6th ed. Brooks/Cole. USA.
13. Wang, Y., Sohn, M. D., Gadgil, A. J., Wang, Y., Lask, K. M., & Kirchstetter, T. W. (2013). How many replicate tests do I need ? – Variability of cookstove performance and emissions has implications for obtaining useful results. Lawrence Berkeley National Laboratory.