Eurasian Journal of Educational Research
www.ejer.com.tr

Applying Bootstrap Resampling to Compute Confidence Intervals for Various Statistics with R

C. Deha DOGAN[1]

A R T I C L E I N F O	A B S T R A C T
	Background: Most of the studies in academic journals use p values to represent statistical significance. However, this is not a good indicator of practical significance. Although confidence intervals provide information about the precision of point estimation, they are, unfortunately, rarely used. The infrequent use of confidence intervals might be due to estimation difficulties for some statistics. The bootstrap method enables researchers to calculate confidence intervals for any statistics. Bootstrap resampling is an effective method of computing confidence intervals for nearly any estimate, but it is not very commonly used. This may be because this method is not well known or people may think that it is complex to calculate. On the other hand, researchers may not be familiar with R and be unable to write proper codes.

Purpose: The purpose of this study is to present the steps in the bootstrap resampling method to calculate confidence intervals using R. It is aimed toward guiding graduate students and researchers who wish to implement this method. Computations of bootstrapped confidence interval for mean, median and Cronbach's alpha coefficients were explained with the R syntax step-by-step. Moreover, traditional and bootstrapped confidence intervals and bootstrapped methods were compared in order to guide researchers. **Main Argument and Conclusions**: With the help of statistical software today it is easy to compute confidence intervals for almost any statistics of interest. In this study R syntax were used as an example so that beginners can use R to compute confidence intervals. Results showed that traditional and bootstrapped confidence intervals have very similar results for normally distributed data sets. Moreover different bootstrapped methods produce different results with skewed data sets. This is because bias corrected and accelerated interval methods are suggested for use with skewed data sets. **Implications for Research and Practice**: R codes presented in this study guide researchers and graduate students while computing bootstrap confidence intervals. Furthermore findings about the comparison of bootstrap methods help researchers choose the most appropriate bootstrap methods. Results and the main argument of this study may encourage researchers to compute bootstrap confidence intervals in their studies.

---

[1] Ankara University, Faculty of Educational Science, Department of Measurement and Evaluation, dehadogan@gmail.com

Introduction

The p value is the probability under the assumption that there is no true effect or no true difference of collecting data that shows a difference equal to or more extreme than what it is actually observed (Reinhart, 2015). Hypothesis testing uses p value to get statistical significance. It is the most widely reported statistic in academic journals. M. Marshall et al. (2000) highlighted that significance of testing is reported in 97% of the research papers in experimental psychology research journals. However the p value has some limitations and it is not an indicator of practical significance. It is well known that confidence intervals provide more information than p values (Haukas & Lewis, 2005). Effect sizes, confidence intervals, and confidence intervals of effect sizes are indicators of practical significance (Banjanovic & Osborne, 2015). Editors of many scientific journals require the use of confidence intervals (Cooper, Wears & Schriger, 2003). Moreover APA's Publication Manual (2001) highlighted the importance of calculating and reporting confidence intervals and effect sizes in academic research.

Unfortunately confidence intervals are rarely reported in academic papers. Reinhart (2015) stated the reasons for this as follows:

> *It's best to do statistics the same way everyone else does, or else the reviewers might reject your paper. Or maybe the widespread confusion about p values obscures the benefits of confidence intervals. Or the overemphasis on hypothesis testing in statistics courses means most scientists don't know how to calculate and use confidence intervals.*

According to Banjanovic and Osborne (2015) the infrequent use of confidence intervals is due to estimation difficulties for some statistics. Some statistics may require multi-step formulas with assumptions that might not always be viable for calculating confidence intervals.

The bootstrap method enables researchers to calculate confidence intervals for any statistics regardless of the data's underlying distribution. The empirical bootstrap was introduced in 1979 (Efron 1988), but it was feasible to implement it without modern computing power. However, computers and statistical software have improved a lot, and today it is possible to calculate confidence intervals using the bootstrap method. Moreover free and open source R software enables researchers to write their own syntax to calculate confidence intervals for various statistics.

Bootstrapping

Briefly, bootstrap methods are resampling techniques for assessing uncertainty. In a broad sense the bootstrap is a widely applicable and extremely powerful statistical tool that can be used to quantify the uncertainty associated with a given estimator or statistical learning method (James et al., 2014). Bootstrap resampling is a method of computing confidence intervals for nearly any estimate. In most studies

the researchers begin with the population and take a sample from the population and run an analysis on that sample. In bootstrap resampling additional sub-samplings and replications are implemented on the original sample. In other words in the beginning of the process thousands of "bootstrapped resamples" are generated from the original sampling using random sampling with replacements. Then the designated statistic (mean, median, regression, Cronbach's alpha coefficient, etc.) is replicated in each of these resamples. Therefore, researchers may get thousands of estimates on the designated statistics. Distribution of those estimates is called "bootstrap distributions". The bootstrap distribution may be used to estimate more robust empirical confidence intervals. In bootstrap sampling the number of replications is very important. Diciccio and Efron (1996) highlight the importance of using at least 2000 replications while conducting bootstrap resampling. A schematic description of the steps for estimating confidence intervals using bootstrap formed by Haukoos & Lewis (2005) is shown in Figure 1.
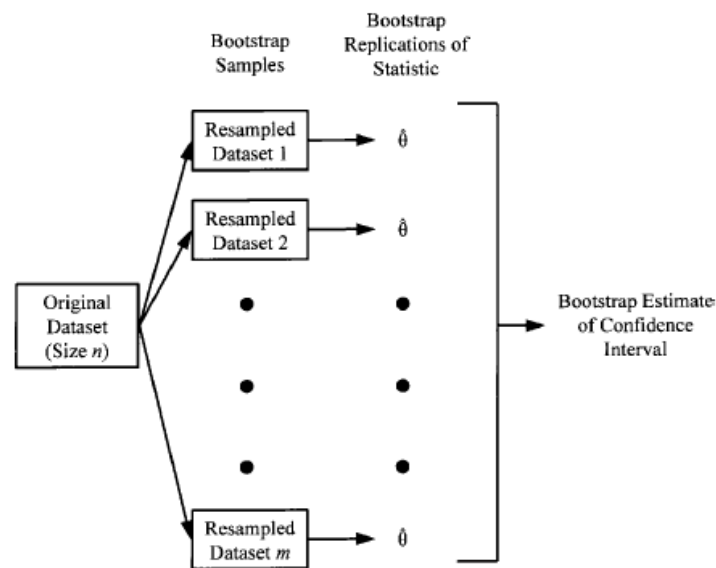

Figure 1: Description of the steps in bootstrapping.

Methods of Bootstrapping

There are different methods for estimating confidence intervals from a bootstrapped distribution. The most frequently used methods are:

• The normal interval method

- The percentile interval method

- The basic interval method

- The bias corrected and accelerated interval method.

The normal interval method computes an approximate standard error using bootstrap distribution (sampling distributions resulting from bootstrap resamples). Then Z distribution is used to get the confidence interval. The percentile interval method uses a frequency histogram of the specific statistics computed from bootstrap samplings. The 2.5 and 97.5 percentiles constitute the limits of 95% confidence intervals (Haukoos & Lewis, 2005). The percentile interval method makes no adjustment while the student interval method corrects each statistic by its associated standard error and converts the distribution to studentized distribution. Then the confidence intervals are found at the 0.025 and 0.975 quantiles as done in the percentile interval method. The bias corrected and accelerated interval method corrects the distribution for bias and acceleration. This method adjusts the distribution based on two coefficient called "bias correction" and "acceleration". The bias correction adjusts for the skewness in bootstrap distribution; it will be zero when bootstrap sampling is perfectly symmetric. On the other hand, coefficients of acceleration do corrections for non-constant variances within the resampled data set (Efron, 1988). Then confidence intervals are found at the 0.025 and 0.975 quantiles of the corrected distribution. The basic interval method corrects the distribution for bias and detects the lower and upper bounds which cover the desired confidence interval (Banjanovic & Osborne, 2015). Each bootstrapping method has advantages and disadvantages, and it is important to use the more appropriate method when computing confidence intervals for the statistic of interest.

Unfortunately, it is not very common to use the bootstrap method to calculate confidence intervals. This may be because they are not well known or people may think that it is complex to calculate. There is statistical software that enables users to compute confidence intervals using bootstrap methods. R is one such software. It is a language and environment for statistical computing and graphics. Because it is free and open-sourced, R has become popular recently for statistical data analysis. Moreover R syntax may seem complicated for people and this may dissuade them from use bootstrapping methods to calculate confidence intervals. Furthermore, comparisons of bootstrapping methods may guide researchers while deciding the methods to be used.

The purpose of this study is to present the steps in the bootstrap resampling method to calculate confidence intervals using R syntax. It is aimed to guide graduate students and researchers who wish to implement bootstrap resampling using R programming language. Computation of bootstrapped confidence intervals for mean, median and Cronbach's alpha coefficient were explained step-by-step using the R syntax. Moreover, some comparisons have been made. Traditional and bootstrapped confidence intervals were compared while computing mean for normally distributed data and median for normally distributed and skewed data sets.

In this study R codes were written on R consoles and directly copied and pasted on this manuscript. This is why R codes have different letter characters and sizes from the other texts in the paper; so many R codes were presented in the paper, they were not defined as figures.

Traditional Confidence Interval for Mean

Traditionally confidence intervals are computed using the formula $\bar{x} \pm 1.96 * (\frac{s}{\sqrt{n}})$ where $\bar{x}$ is the mean and s is the standard deviation and n is the sample size. In this part at first a pedagogical example is presented to compare traditional confidence intervals and bootstrap confidence intervals. We will first generate random data with 100,000 observations. This data set is viewed as population. The population has random normal distribution (Mu=60, Sigma=7) with 100,000 observations.

```
> population<-rnorm(100000,60,7)
> str(population)
 num [1:100000] 64.5 65.2 51.9 61.5 50.5 ...
```

Next we take six random samples of 50 observations from the population in order to compute traditional confidence intervals.

```
> list<-vector("list",6)
> for ( i in 1:6) {list[[i]]<-sample(population,50)}
> str(list)
List of 6
 $ : num [1:50] 51.3 63.5 74.7 60.5 62.8 ...
 $ : num [1:50] 57.5 48.7 70.2 74.8 57 ...
 $ : num [1:50] 52.9 50.3 63.2 56.7 62.3 ...
 $ : num [1:50] 57 55.3 62.7 69.4 62 ...
 $ : num [1:50] 59.5 56.6 59.1 55.6 58.7 ...
 $ : num [1:50] 59.9 54.4 52.8 57.5 52.5 ...
```

The next step is to write a simple function to calculate confidence intervals for the six samples taken randomly from the population.

```
> conf.interval<-function(x){ mean=mean(x); sd<-sd(x); n<-length(x)
+ se<-1.96*(sd/(n**.5)); ci<-c(mean=mean,lowerci=mean-se,upperci=mean+se)
+ return(ci)}
> conf.interval(list[[1]])
    mean  lowerci  upperci
59.85779 57.66439 62.05119
```

Now it is easy to compute 95% confidence intervals for the samples randomly taken from the population. We can round the results to two digits using the round () function.

```
> round(conf.interval(list[[1]]),2)
   mean lowerci upperci
  59.86   57.66   62.05
```

We can also write another function to take six random samples from the population. It would be better to define set.seed () for each sample. We will use the same samples to generate six different bootstrap distributions so we can compare confidence intervals.

```
> conf.interval1
function(x,seed) { set.seed(seed)
sample<-sample(x,50); mean<-mean(sample);sd<-sd(sample)
n<-length(sample); se<-1.96*(sd/(n**.5))
ci<-round(c(mean=mean,lowerci=mean-se,upperci=mean+se),2)
return(ci)}

conf.interval1(population,seed=10)
 mean lowerci upperci
61.33   59.76   62.90
conf.interval1(population,seed=20)
 mean lowerci upperci
60.46   58.27   62.65
conf.interval1(population,seed=30)
 mean lowerci upperci
59.73   57.55   61.91
conf.interval1(population,seed=40)
 mean lowerci upperci
59.21   57.33   61.09
conf.interval1(population,seed=50)
 mean lowerci upperci
61.71   59.53   63.89
conf.interval1(population,seed=60)
 mean lowerci upperci
61.24   59.03   63.45
```

Now we can compute 95% confidence intervals of six samples randomly taken from the population. We defined set.seeds as 10, 20, 30, 40, 50 and 60 for 6 samples respectively.

Bootstrapped Confidence Interval for Mean

In order to get bootstrap distribution, the "boot" package will be used (Canty & Ripley 2016). The "boot" function is used to generate bootstrap distribution for specific samples, but this function requires writing simple functions about the statistics of interest.

```
> select
function(x,seed) { set.seed(seed)
sample<-sample(x,50)
return(sample)}
> mean.func
function(x,i){ mean(x[i]) }
```

The first function is used for the first argument of "boot" function. It is used to take the specific sample from the population. The second one is the statistics of interest that is "mean" in this example.

```
> boot.sample1<-boot(select(population,seed=10),mean.func,R=2000)
> boot.sample2<-boot(select(population,seed=20),mean.func,R=2000)
> boot.sample3<-boot(select(population,seed=30),mean.func,R=2000)
> boot.sample4<-boot(select(population,seed=40),mean.func,R=2000)
> boot.sample5<-boot(select(population,seed=50),mean.func,R=2000)
> boot.sample6<-boot(select(population,seed=60),mean.func,R=2000)
```

So we generated six bootstrap distributions with 2,000 resamples. We used the "select" and "mean.func" functions within the "boot" functions. Below is the bootstrap statistics for the first sample.

```
Bootstrap Statistics :
    original      bias    std. error
t1* 61.33011 -0.02389115   0.8002305
```

The value original is the mean of the whole sample while bias is the difference between the original mean and the mean of the bootstrapped samples. Standard error is the standard deviation of the simulated values. The next step is calculating confidence intervals for each original and bootstrapped samples. We used the "boot.ci" function to compute confidence intervals for each bootstrapped sample.

```
> ci.boot1<-boot.ci(boot.sample1) > ci.boot2<-boot.ci(boot.sample2)

> ci.boot3<-boot.ci(boot.sample3) > ci.boot4<-boot.ci(boot.sample4)

> ci.boot5<-boot.ci(boot.sample5) > ci.boot6<-boot.ci(boot.sample6)
```

```
Intervals :
Level      Normal              Basic
95%   (59.79, 62.92 )    (59.75, 62.94 )

Level     Percentile            BCa
95%   (59.72, 62.91 )    (59.79, 62.96 )
Calculations and Intervals on Original Scale
```

The "basic" title in the R output refers to "basic interval method". In this method confidence intervals are estimated by correcting the bootstrap distribution for bias or skew. The "Bca" title in R output refers to the "bias corrected and accelerated interval method". In the Bca method the bootstrap distribution is corrected for bias and

acceleration and the CI are found at the .025 and .975 quantiles of the corrected distribution (Carpenter & Bithell 2000; Banjanovic & Osborne, 2015).

Comparison of Traditional and Bootstrapped Confidence Intervals for Mean

The result of the "boot.ci" function produces confidence intervals for four main bootstrapped methods discussed earlier. Table 1 presents the comparison of confidence intervals of each original random sample and its associated bootstrap distributions.

Table 1

Confidence Intervals for Original and Bootstrapped Samples

Sample	Mean	Traditional 95% CI		Bootstrapped 95% CI	
		Lower Bound	Upper Bound	Lower Bound	Upper Bound
1	61.33	59.76	62.90	59.79	62.92
2	60.46	58.27	62.65	58.34	62.59
3	59.73	57.55	61.91	57.54	61.94
4	59.21	57.33	61.09	57.32	61.08
5	61.71	59.63	63.89	59.53	63.88
6	61.24	59.03	63.45	58.97	63.46

As seen in Table 1 two methods produce very similar results. The differences are at the first or second decimal place. However, it is very important to note that those methods will yield very similar results unless the data violates parametric assumptions such as normality (Banjanovic & Osborne, 2015).

Bootstrapped Confidence Intervals for Median

Median is the observation at the 50th percentile in a set of data ordered from the lowest value to the highest value. It is commonly reported and considered a more valid definition of center when the frequency distribution of the variable is skewed. No simple formula exists for computing confidence intervals for median. According to central limit theorem, the number of resampled data sets increasing the distribution of the resulting statistic will become approximately normal (Zar, 1999). However, using the bootstrapped resampling method, it is possible to calculate the confidence interval for median.

In this example we will calculate the confidence interval for median using two different data sets; one is normally distributed and the other is skewed. Next we will compare bootstrapped methods with each other.

We had a normally distributed data set with 1,000,000 observations called "population". Now we generate another data set called "population 2" with

1,000,000 observations using skewed chi-square distribution. In the previous example the population had 100,000 observations. There is no technical reason for increasing the number of the observations from 100,000 to 1,000,000. Both are large enough. It is aimed to demonstrate to the reader how R could easily generate big data sets.

```
> population2<-rchisq(1000000,5)
> str(population2)
 num [1:1000000] 1.82 2.05 2.62 3.71 3.22 ...
> |
```

First we will compute a confidence interval for median using a normally distributed population. We compute bootstrapped distributions and bootstrapped confidence intervals using the functions below, previously mentioned in detail. Additionally we added a new simple function called "median.func" to calculate the median within the "boot" function.

```
> median.func
function(x,i) { median(x[i]) }

> norm.sample<-boot(select(population,seed=15),median.func,R=2000)
> norm.sample


> boot.ci(norm.sample)
```

Then we will compute bootstrapped confidence intervals for the median using skewed distributed population.

```
> skew.sample<-boot(select(population2,seed=15),median.func,R=2000)

> boot.ci(skew.sample)
```

Comparison of Bootstrapping Methods Regarding Distributions of Data Sets

Table 2 summarizes the confidence intervals for normally distributed and skewed data sets regarding bootstrapping methods.

Table 2

Comparison Confidence Intervals Regarding Bootstrapping Methods

Methods for Bootstrapping	Normally Distributed Data		Skewed Data	
	95% Confidence Interval		95% Confidence Interval	
	Lower Bound	Upper Bound	Lower Bound	Upper Bound
The normal interval method	56.80	59.04	4.133	6.071
The percentile interval method	56.87	59.35	4.081	5.865
The basic interval method	56.66	59.14	4.295	6.078
Bias corrected and accelerated interval method	56.86	59.32	3.740	5.848

Table 2 shows that all bootstrapping methods for normally distributed data sets produce very similar confidence intervals. The only differences are at the first or second decimal place. On the other hand, for skewed data sets bootstrapping methods produce different confidence intervals. Therefore, different methods of skewed data sets of bootstrapping may produce different results. Before deciding the bootstrapping method to be used, their assumptions should be taken into consideration. The bias corrected and accelerated interval method requires no assumptions about the distribution of the data sets while others do. So it is better to consider using this method to compute confidence intervals with skewed data set.

Bootstrapped Confidence Intervals for Cronbach's Alpha Coefficient

Cronbach's alpha coefficient is an indicator of reliability that is commonly used, especially in psychological tests. In fact, it is the indicator of internal consistency. Many researchers use Cronbach's alpha coefficient for a set of items to construct a scale. Coefficient alpha (commonly called Cronbach's alpha) was developed by Lee Cronbach in 1951 to provide a measure of the internal consistency of a test or scale; it is expressed as a number between 0 and 1. Calculating alpha has become common practice because it is easier to use than other estimates (e.g. test retest reliability estimates) as it only requires one test administration (Tavakol & Dennick, 2011). Item-total correlations are the correlation between an item and all other items, where the total of the other items is achieved by summing and averaging them (Banjanovic & Osborne, 2015). Moreover computing confidence intervals for Cronbach's alpha coefficient and item total correlations provides a very good indication of the generalizability of the results.

To compute traditional item total correlations and Cronbach's alpha coefficient the "CTT" package was used (Willse, 2014). In order to compute bootstrapped confidence intervals for Cronbach's alpha coefficient and item-total correlations, "boot" package has been used. The data is taken from a five-item scale that was administered to 300 hundred participants.

First the item total correlations and Cronbach's alpha coefficient were computed with the codes given below. "cr.data1" is the original data set. The object "cr.alp" was created to compute the statistics of interest. "cr.alp$alpha" is the Cronbach's alpha coefficient and "cr.alp$pBis" is the item total correlations for each of the five items.

```
> cr.alp<-reliability(cr.data1,itemal=TRUE)

> cr.alp$alpha
[1] 0.8346358
> cr.alp$pBis
[1] 0.5964466 0.7006749 0.6642530 0.6159572 0.6021944
```

Before bootstrapping 2,000 resamples we were required to write a simple function to use within the boot function. "cr.samp1" is the function that might be used to compute Cronbach's alpha coefficients and item total correlations for bootstrapped resamples.

```
> cr.samp1
function(data, indices) {
sample<-data[indices,]
cr.alpha<-reliability(sample, itemal = TRUE)
result=with(cr.alpha,c(alpha,pBis))
return(result)}
```

Next 2,000 bootstrapped resamples were generated using the boot function given below. This process may take 30 minutes to 1 hour depending on the size of the original sample and the number of the bootstrapped samples. The values named as original in the output are the Cronbach's alpha coefficient (the first row) and item total correlations (rows 2 to 5) for the original data set. Bias is the difference between the original and the bootstrapped values. Standard error is the standard deviation of the simulated values.

```
> cronbach <- boot(data=cr.data1, statistic=cr.samp1, R=2000)
```

```
Bootstrap Statistics :
     original        bias     std. error
t1* 0.8346358 -0.0014286370  0.02290934
t2* 0.5964466  0.0017199229  0.05168665
t3* 0.7006749 -0.0001259864  0.03830456
t4* 0.6642530 -0.0016074008  0.04738270
t5* 0.6159572 -0.0007566143  0.05757076
t6* 0.6021944  0.0001811386  0.05587665
```

The next task is to compute confidence intervals for Cronbach's alpha coefficient and item total correlations. The function "boot.ci" was used to calculate those intervals, but we needed to write another function including the "boot.ci" function. This function is called "computeCI" When writing the code, I was inspired from the code written by Banjanovic and Osborne (2015). The first line in the output gives confidence intervals for Cronbach's alpha coefficient and the rest of the lines give the intervals for item total correlations for each item respectively. I restricted the output only with the "percentile" method of bootstrapping discussed above.

```
> computeCI<-function(alpha){
+ Conf.Interval<-do.call("rbind", sapply(1:6, FUN = function(i)
+ boot.ci(alpha,type="perc", index=i)$perc[4:5], simplify = FALSE))
+ return(Conf.Interval)}
> computeCI(cronbach)
          [,1]       [,2]
[1,] 0.7851932 0.8749758
[2,] 0.4932083 0.7001377
[3,] 0.6206685 0.7704308
[4,] 0.5632592 0.7518773
[5,] 0.4953220 0.7198234
[6,] 0.4808203 0.7054129
```

Table 3 presents the values and confidence intervals for the statistics of interest.

Table 3

Item Total Correlations, Cronbach's Alpha Coefficient, and Related Confidence Intervals

Items	Original Item Total Correlation	Bootstrapped confidence Intervals	
		Lower CI	Upper CI
1	0.60	0.49	0.70
2	0.70	0.62	0.77
3	0.66	0.56	0.75
4	0.62	0.49	0.71
5	0.60	0.48	0.71
Cronbach Alpha	0.83	0.79	0.87

As seen in Table 3, Cronbach's alpha coefficient for the scale is 0.83 and the item total correlations vary between 0.60 and 0.70. The confidence intervals for Cronbach's alpha is relatively small compared to those for item total correlations (0.79 - 0.87). On the other hand, confidence intervals for item total correlations are wide. This may lead us to expect that there will be significant variability in the point to estimate for item total correlations in other similar samples. The purpose of this study is not to construct or adopt a scale, so the results are not discussed in detail. But the point is that computing confidence intervals for Cronbach's alpha coefficient and item total correlations may guide the researcher while constructing scales.

Discussion

Confidence intervals are good indicators of practical significance, unlike p values and they also provide more information than p values (Haukas & Lewis, 2005). Unfortunately, confidence intervals are rarely reported in academic papers. This is because computing confidence intervals are not practical and not possible for some statistics. This is why bootstraps methods, which are resampling techniques for assessing uncertainty, have become popular.

In this study the basic principles of bootstrapping and some commonly used bootstrapping methods were briefly presented. Then computation of bootstrapping methods for mean, median and Cronbach's alpha coefficient were explained using R syntax. Furthermore, some comparisons were done. Traditional and bootstrapped confidence intervals were compared for mean. Moreover bootstrapped methods were compared while calculating median for normally distributed and skewed data sets.

With the help of statistical software today it is easy to compute confidence intervals for almost any statistics of interest. R, which is a free and open-sourced software, is one of them. This is why in this study R syntax was used as an example so that beginners could use the R software to compute confidence intervals. Those syntax are not the only solutions but they are just some of the many other possible ways to write syntax while computing confidence intervals.

The results of the comparisons done in this study show that with normally distributed data, traditional and bootstrapped confidence intervals are close to each other. However, with skewed data sets this may not be so. In these cases the bootstrapped methods come into prominence.

The other important point is deciding the bootstrapped methods to be used when computing confidence intervals. Banjanovic and Osborne (2015) present very effective strategy to decide the bootstrapped methods to be used. They suggest that researchers answer four simple questions.

1.   Is there a formula to estimate the standard error of the statistics?

2. Is the distribution symmetrical around the mean of the bootstrapped distribution?

3. Is the distribution normal?

4. Is the sample estimate a biased estimate of the population statistics?

The answer for the first question is "yes" for mean but "no" for median. For example, for small samples or skewed distributions it may be advised to run different methods. If the results differ from each other it may be advised to use the bias corrected and accelerated interval method since it has no assumptions about the distributions while the others do. Therefore, for the example presented in Table 2 it would be better to use the bias corrected and accelerated interval method. However, Banjanovic and Osborne (2015) highlighted that bootstrap analysis may not repair highly biased samples. This principle should be kept in mind. For in-depth information about the assumptions of bootstrapped methods, see Davison & Hinkley (1997), Carpenter & Bithell (2000) and Chernick & Labudde (2011).

Using bootstrapped analysis is also very effective for statistics that already have a formula to calculate standard error. Although it is possible to compute standard errors for intercept and slope of the regression line in linear regression analysis, bootstrapped analysis may produce more accurate standard errors (James et al. 2014).

In this study computing bootstrapped methods for mean, median and Cronbach's alpha coefficient were illustrated because they are commonly used in social sciences. But it is also possible to compute bootstrapped confidence intervals for almost any statistics of interest such as effect size, linear regression, logistic regression, factor analysis, etc. Each of them may be calculated using "boot" packages illustrated in this paper. But this package requires writing a simple chunk of code about the statistic of interest, which can be used within the "boot" function.

Briefly, with help of statistical software like R it is possible to conduct bootstrapped analysis. Researchers are now able to compute confidence intervals for almost any statistics. Unfortunately, reporting bootstrapped confidence intervals in academic journals is very rare. In this study basic steps for calculating bootstrapped confidence intervals for some commonly used statistics were illustrated using R syntax. Therefore, it is expected that this study guide will enable researchers to compute bootstrapped confidence intervals and lead them to report confidence intervals in their academic studies.

References

Angelo Canty and Brian Ripley (2016). boot: Bootstrap R (S-Plus)   Functions. R package version 1.3-18

American Psychological Association. (2001). *Publication manual of the American Psychological Association*, 5th Edition. Washington, D.C.: American Psychological Association.

Banjanovic, Erin S., and Osborne, Jason W. (2016). Confidence Intervals for Effect Sizes: Applying Bootstrap Resampling. *Practical* Assessment, *Research & Evaluation*, 21(5).

Carpenter, J. and Bithell, J. (2000), Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians. *Statistics in Medicine*, 19: 1141–1164.

Cooper RJ, Wears RL, Schriger DL (2003) Reporting research results: recommendations for improving communication. *Ann Emerg Med*. 41:561–4.

Chernick, M. R., & Labudde, R. A. (2011). *An introduction to bootstrap methods with applications to R*. A john Wiley & Sons, Inc. New Jersey.

Davison, A. C., & Hinkley, D. V. (1997). *Bootstrap methods and their application. Cambridge*, United Kingdom: Cambridge University Press.

DiCiccio, T. J., & Efron, B. (1996). Bootstrap confidence intervals. Statistical science. *A Review Journal Of The Institute Of Mathematical Statistics.*, 11(3), 189-212.

Efron, B. (1988). Computer intensive methods in statistical regression. *Slam Review*. 30 421-449

Haukoos, J. S., Roger, Lewis, R. J. (2005). Advanced statistics: Bootstrapped confidence interval for statistics with "difficult" distributions. *Academic Emergency Medicine* 12(4), 360-354.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2014). *An introduction to statistical learning: with applications in* R. Springer, New York. USA

M. Marshall, A. Lockwood, C. Bradley, C. Adams, C. Joy, and M. Fenton. (2000). Unpublished rating scales: a major source of bias in randomized controlled trials of treatments for schizophrenia. *The British Journal of Psychiatry* 176 (3) 249–252.

Reinhart, A., (2015). *Statistics done wrong. The woefully complete guide*. No Stretch Press. San Francisco, USA.

Tavakol, M., Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*. 2:53-55.

John T. Willse (2014). CTT: Classical Test Theory Functions. R package version 2.1. https://CRAN.R-project.org/package=CTT

Zar JH. (1999). Biostatistical analysis.4th edition. Englewood Cliffs, NJ: Prentice Hall.

Bootstrap Örnekleme Yönteminin Çeşitli İstatistikler için Güven Aralığının Hesaplanmasında R Yazılımı ile Kullanımı

Çalışmanın Temeli: Yokluk hipotezine dayalı istatistiksel testler istatistiksel manidarlığı belirlemek için p değerini kullanır. Akademik dergilerde en çok rapor edilen istatistik olan p değeri pratik manidarlık anlamında önemli sınırlılıklara sahiptir. Güven aralığının p değerine kıyasla daha fazla bilgi sağladığı belirtilebilir. Etki büyüklüğü, güven aralığı ve etki büyüklüğüne ilişkin güven aralığı pratik manidarlığın göstergeleri olarak görülür. Pek çok akademik dergi editörü güven aralığı hesaplarının makalelerde rapor edilmesi gerektiğini vurgulamaktadır. Aynı şekilde APA 2001 raporunda güven aralığı ve etki büyüklüğünün raporlanmasının önemi vurgulanmıştır. Ancak tüm bunlara rağmen akademik makalelerde güven aralığının raporlanma oranı çok düşüktür. Bunun bir nedeni araştırmacıların çok yaygın rapor edilen p değerinden daha farklı bir istatistiği rapor etme konusundaki çekinceleri olabilir. Bu durum araştırmacıların hatalı da olsa grubun büyük çoğunluğuna uyum sağlama davranışı sergilemelerine ve bu sayede makalenin kabul edilme olasılığına artırmak istemelerine dayalı olabilir. Bir diğer neden ise bazı istatistikler için güven aralığının hesaplanmasının pratik veya mümkün olmamasıdır. Bootstrap örnekleme yöntemi güven aralığı hesaplama sürecinde kullanılabilecek güçlü bir istatistiksel yaklaşımdır. Temel amaç evrenden seçilen bir örneklem üzerinden yerine koyarak örnekleme yöntemi ile yeni örneklemler oluşturmaya dayanır. Elde edilen yeni örneklemlere ilişkin istatistikler hesaplanarak standart hata ve güven aralığı kestirimi yapılabilir. Bu sayede geleneksel olarak standart hata ve güven aralığı tahmini yapılamayan pek çok istatistik için güven aralığının hesaplanmasına olanak tanır. Yerine koyarak örnekleme yöntemi kullanıldığı için orijinal örneklemden daha fazla gözlem sayısını içeren örneklemler oluşturulabilir. Temelleri 1979 yılında atılan bootstrap analizinin kullanımı bilgisayar yazılımlarının gelişimi ile yaygınlık kazanmıştır. Yazılımların da yardımı ile pratik bir şekilde hesaplanabilecek bu yöntem maalesef çok nadir kullanılmakta ve araştırma makalelerinde rapor edilmemektedir. Bu durumun temel nedeni bootstrap yönteminin ve bu süreçte R yazılımının nasıl kullanılacağının yeteri kadar bilinmemesi yatabilir. Alan yazına bakıldığında bootstrap yönteminin R kodları ile açıklandığı çalışmaların sınırlı olduğu belirlenmiştir.

Çalışmanın Amacı: Bu çalışmada bootstrap örnekleme yöntemine dayalı olarak yaygın kullanılan bazı istatistikler için güven aralığı hesaplama sürecinin R kodları ile açıklanması amaçlanmıştır. Çalışmada öncelikli olarak bootstrap örnekleme yöntemine ilişkin kuramsal bilgiler verilmiş ve kullanılabilecek farklı bootstrap yöntemlerine ilişkin açıklamalar yapılmıştır. Akabinde sosyal bilimlerde yaygın olarak kullanılan aritmetik ortalama, ortanca ve Cronbach Alfa içtutarlık katsayısına ilişkin bootstrap güven aralığı hesaplama süreci örneklendirilmiştir. İlk olarak R kodları kullanılarak aritmetik ortalama için geleneksel güven aralıkları ve "boot" paketinden faydalanarak bootstrap güven aralıkları hesaplanmış ve sonuçlar karşılaştırılmıştır. Daha sonra normal ve çarpık olan iki dağılım üzerinden ortanca için bootstrap güven aralıkları hesaplanmış ve farklı bootstrap yöntemleri bu

bağlamda karşılaştırılmıştır. Son olarak Cronbach Alfa içtutarlık katsayısına ilişkin bootstrap güven aralıkları hesaplanmış ve örneklendirilmiştir.

Kaynak Kanıtları: Çalışma sürecinde geleneksel güven aralığı hesaplama sürecine ilişkin R kodları sunulmuştur. Ayrıca "boot" paketinden faydalanılmış ve bu paket içerisinde yer alan "boot" ve "boot.ci" fonksiyonları açıklanmıştır. Bu paket içerisinde yer alan "boot" ve "boot.ci" fonksiyonları güven aralığı için hesaplanacak her istatistik için yeni kodların yazımını gerektirmektedir. Bu kodlar da araştırma içerisinde sunulmuş ve nasıl kullanılacağına yönelik açıklamalarda bulunulmuştur. Çalışmanın özet bölümün sayfa sınırlılığı nedeni ile orijinal İngilizce metninde sunulan R kodları ve fonksiyonlar burada verilmemiştir. Ancak tartışma ve sonuçlar bölümünde öne çıkan bulgular ve ilgili yöntemin kullanımına ilişkin önemli açıklamalar sunulmuştur.

Tartışma ve Sonuçlar: Bootstrap örnekleme yöntemi uygulama ve yorumlama açısından oldukça pratiktir. R programlama diline ilişkin temel düzeyde bilgiler ile bootstrap örnekleme yöntemi kolaylıkla gerçekleştirilebilir. Çalışma kapsamında gerçekleştirilen karşılaştırmada aritmetik ortalama için geleneksel ve bootstrap yöntemi ile kestirilen güven aralıklarının birbirine çok yakın olduğu gözlenmiştir. Ancak bu durumun normal dağılıma sahip veri setleri için böyle olduğu belirtilebilir. Bunun yanı sıra çarpık ve normal dağılıma sahip veri setleri üzerinden ortanca için kestirilen bootstrap güven aralıklarının farklı yöntemlere göre değiştiği görülmüştür (Normal aralık, yüzdelikli aralık, temel aralık ve yanlılık düzeltmeli aralık). Çarpık dağılımlar için "yanlılık düzeltmeli aralık" yönteminin daha doğru sonuç verdiği alan yazında vurgulanmıştır. Ölçek geliştirme sürecinde kullanılan gerçek veri seti üzerinde Cronbach Alfa içtutarlık katsayısı ve madde toplam korelasyonları için hesaplanan bootstrap güven aralıklarının ise oldukça geniş olduğu gözlenmiştir. Bu bağlamda ölçek uyarlama ve geliştirme sürecinde bootstrap güven aralıklarının hesaplanmasının araştırmacılara önemli boyutta yol göstereceği ve daha nitelikli araçların geliştirilmesine katkı sağlayacağı belirtilebilir. Bootstrap örnekleme yönteminin uygulaması pratik olmakla beraber önemli kuramsal temellere dayandığı belirtilebilir. Bu bağlamda ilgili okuyucuların İngilizce tam metinde önerilen kaynaklara ulaşmaları önerilir. Bu çalışmanın araştırmacılara bootstrap örnekleme yöntemini kullanma sürecinde yol göstermesi ve araştırmacıları güven aralıklarını rapor etmeleri konusunda teşvik etmesi beklenmektedir.