Available online at www.sciencedirect.com

SciVerse ScienceDirect

journal homepage: www.elsevier.com/locate/cegh

Original Article

Confidence Interval for skewed distribution in outcome of change or difference between methods

J. Visalakshi^a, L. Jeyaseelan^{b,*}^a Lecturer, Department of Biostatistics, Christian Medical College, Vellore 632 002, India^b Professor, Department of Biostatistics, Christian Medical College, Bagayam, Vellore 632 002, India

ARTICLE INFO

Article history:

Received 15 May 2013

Accepted 18 July 2013

Available online 20 August 2013

Keywords:

Standard error

Confidence Interval

Bootstrap

Skewed

Log transformation

ABSTRACT

Background: When we have pre and post measurements on same subjects and the outcome of interest is change or to compare the reliability across two methods, then it is required to present mean change and the 95% Confidence Interval (CI) for the change. However, when the distribution of the 'change' is skewed, then it is not possible to calculate CI using normal approximation. This study was to demonstrate an appropriate method in such situations.

Methods: Hypothetical data was considered. Difference of two methods was obtained that included positive and negative values and 95% CI using normal approximation with log transformation, Hodges–Lehmann CI, shifting the origin with log transformation and the Bootstrap CI was obtained.

Results: Data consisted of 399 observations. The mean (sd) of the outcome was 96.9 (465.6). The 95% CI using the normal approximation with log transformation was obtained as (245.8, 307.5) with only 194 observations while Bootstrap CI was calculated as (54.1, 139.4) using all observations.

Conclusion: When the outcome of interest is to compare the 'change' which is skewed, then we discourage the log transformed normal approximation method or adding constant and taking log transformation method to calculate CI and encourage researchers to use Bootstrap CIs.

Copyright © 2013, INDIACLEN. Publishing Services by Reed Elsevier India Pvt Ltd. All rights reserved.

1. Introduction

Very often in clinical trials, the main objective of the study is to compare the change before (pre) and after (post) intervention. Also, in some reliability studies, it is of interest to see if two methods are significantly different when measuring the same quantity. In such situations, the main interest is to estimate the amount of change and the likely variability in the long run (95% CI). However, in order to test whether the change is significant or not, we examine the 95% CI, whether that CI includes zero or

not. If 'change' is defined as post measurement minus the pre measurement, then there are positive values when the post measurements are greater than the pre and negative values when the pre measurements are greater than post. The only way to test for such a hypothesis is to perform a one sample t-test or a non-parametric Wilcoxon Signed Rank Test for the difference or change. However, one sample t-test is appropriate when the distribution of the difference is approximately normal. When the distribution of the difference or 'change' is positively skewed, then it is recommended to take a log

* Corresponding author. Tel.: +91 0416 2262703, 2284205; fax: +91 0416 2262703.

E-mail addresses: ljey@hotmail.com, ljey@cmcvellore.ac.in (L. Jeyaseelan).

2213-3984/\$ – see front matter Copyright © 2013, INDIACLEN. Publishing Services by Reed Elsevier India Pvt Ltd. All rights reserved.
<http://dx.doi.org/10.1016/j.cegh.2013.07.006>

transformation.^{1,2} But many software packages like SAS, SPSS, STATA, etc. produce missing values for all the negative differences as the log transformation of negative values does not exist. The main objective of the study is to demonstrate the misuse of taking log transformations when the outcome of interest is ‘change’ with positive and negative values and obtain the CI using Bootstrap method in such situations.

2. Methods

2.1. Data

The data used is a hypothetical data of individuals comparing the Total Liver Volume (TLV) (in cubic cm) as was measured from CT scan of the abdomen or thorax and the TLV (in cubic cm) obtained using the Indian formula with the Body Surface Area (BSA), gender and age as the independent variables. The TLV difference between the two methods was obtained, named as “Difference in TLV”, was the outcome of interest. The ‘Difference in TLV’ had both positive and negative values.

2.1.1. Normal approximation for log transformed data

The log transformation of the ‘Difference in TLV’ was obtained and the normal approximation method of calculating 95% CI was obtained.

2.1.2. Change of origin and its log transformation

The minimum value for the ‘Difference in TLV’ was obtained. One positive value that was greater than the minimum value in the data was added to all observations and thereby all observations were only positive in the modified data. The log transformation was taken for the modified data and 95% CI was obtained for log transformed data. Then that 95% CI was back transformed to its original scale (antilog of the CI was taken and the value added was then subtracted from the upper and lower limits of CI) and was interpreted.

2.1.3. Hodges–Lehmann Confidence Interval

A non-parametric method of obtaining CI for median was also calculated. This method was known as Hodges–Lehmann CI for ordered statistics.³ As this method is a non-parametric method, this method is computed using ranks.⁴

2.2. Bootstrap simulation

A random sample of the ‘Difference in TLV’ of all ($n = 399$) observations was obtained using sampling with replacement. ‘Sampling with replacement’ represents that any observation from the dataset can be obtained in the Bootstrap sample more than once. For example, observation 5 in the dataset can occur more than once in the Bootstrap sample and each Bootstrap sample has 399 observations. This process of obtaining the sample was repeated for 10,000 times. After arranging the 10,000 sample means in ascending order, the 2.5th and the 97.5th percentiles provided the 95% CI,⁵ which is generally used when the statistics of interest is median or intra class correlation, ratio measures etc. Usually these statistics have complicated distributions. This method is the Bootstrap Percentile method. R software was used to analyze the data. R software is

an open source software that can be downloaded.⁶ The codes to obtain the Bootstrap percentile CI and Bias Corrected and Accelerated (BCa) CI have been provided in the [Appendix](#). When the distribution of the statistic is skewed, there is a better way of providing CI in bootstrap known as BCa. The BCa provides CI correcting for bias when the data is skewed.⁷

3. Results

The hypothetical data consists of 399 observations. The mean (sd) of the ‘Difference in TLV’ was 96.9 (465.6). [Fig. 1](#) shows the distribution of ‘Difference in TLV’. The figure shows that the data had both positive and negative values and it is quite evident that the distribution is skewed to the right, indicating that data consists of more positive change values. [Table 1](#) shows the CI using normal approximation for log transformed data, CI obtained after changing origin and log transformation, Hodges–Lehmann CI and, the Bootstrap method for ‘Difference in TLV’. The CI using log transformed data, Change in origin and log transformed, Hodges–Lehmann CI was obtained as (245.8, 307.5), (−100.9, −9.5), (−6.5, 80.6) respectively. The Bootstrap BCa CI was found to be (53.2, 147.7). This indicated that there was a clear evidence of significant difference between the two methods ([Table 1](#)). The 95% CI for the mean using the log transformed normal approximation showed an evidence of significance. However, when the log transformed values of the mean difference was considered, the number of observations got reduced to 194 instead of 399. The reason for this is that the log transformation of the negative values is not possible, thereby producing missing values and hence excluded from the analysis. The 95% CI using Hodges–Lehmann ordered statistics was found to be (−6.5, 80.6) suggesting that there was no significant difference between two hypothetical methods. The method that used adding an appropriate constant and taking log transformation provided with the 95% CI which

Fig. 1 – Distribution of difference in TLV.

Table 1 – 95% CI for the difference using different methods.

Variables	n	Mean	sd	95% CI
Log transformed CI for difference in TLV	194	274.94	3.13	245.79, 307.54
CI using the normal approximation	399	96.86	465.60	51.18, 142.55
CI using origin change and log transformation for difference in TLV	399	–57.04	2.23	–100.93, –9.55
CI using Hodges–Lehmann statistics (median and IQR) for difference in TLV	399	–34.63	–222.10, 295.90	–6.52, 80.59
Bootstrap confidence for difference in TLV using Bca method	399	97.16	23.52	53.22, 147.73

Note: CI – Confidence Interval.

suggested statistical significant difference however, the direction of the significance is not in the same direction as it should have been observed.

Bootstrap CIs (BCa) and the normal approximation were almost similar as the sample size was large ($n = 399$). Hence a random sample of different sizes ($n = 15, 20, 25, 30, 35, 40, 50, 60$ and 75) of the 399 observations was taken and it was found that there was a considerable difference in the 95% CI when the sample size was small (when $n < 60$) which is presented in Table 2.

4. Discussion

Confidence Interval for a mean difference is useful to determine the lower and upper limit of that mean difference within which the population parameter lies in a long run. CIs also provide significance of the test. If the hypothesized null value lies between the two limits of the CI, then there is no significant difference between the two groups. If the null value is not contained within the CI, then we conclude that there is a significant difference between the two groups. Apart from the significance of the test, the magnitude of the difference can also be obtained using the lower and the upper limits of the CI. It is a common practice to calculate the CI rather than just reporting p values now-a-days.⁸ When the non-parametric method of calculating CI was used, there was no evidence of significant difference (Table 1). The non-parametric CI produced the 95% CI for the median and not mean. Moreover, the non-parametric method of calculating CI needed an assumption of symmetry.⁹ The log transformation of only positive values due to the shift in the origin was not normal. As a result of this, the CI though suggested significance was in the reverse direction. The

Bootstrap CI suggested that there was a significant difference between the two methods (Table 1). The mean and standard deviation of the difference from the Bootstrap is similar to the mean and sd obtained from the original data. Bootstrap technique does not require any transformation as it takes samples of the same size with replacement. The normal approximation CI ignoring skewness was similar to the Bootstrap CI and the possible reason for the similarity is due to the large number of observations ($n = 399$) in the present study. As Bootstrap technique generates samples from the data with replacement for finitely large number of times, this method can be applied for small sample sized data.^{10–12} Bootstrap method is a general technique that can be applied to many non-standard situations such as regression analysis when the assumptions are violated etc.¹³ Bootstrap method is a quite well recognized method to provide CI for large sample, skewed data and small sample size.¹⁴ Thus, the researchers are expected to calculate CI using Bootstrap. Moreover, CIs are calculated using the standard errors (SE). Standard errors cannot be easily obtained in all situations. It is very difficult to obtain SEs for some estimates like median, ratio measure of two values,¹⁵ etc. Hence the most efficient way to obtain the CI is to obtain the Bootstrap CIs. The data described in the paper is a hypothetical data. Though the data is a hypothetical data, it has been modified in such a way that there are only about 6% of the observations that had differences more than 1000 cubic cm.

Conflicts of interest

All authors have none to declare.

Appendix

```
# Compute 95% CI using already available boot function in R.
# The R codes are case sensitive
# Using the package "boot" in R
# Data: The data can be downloaded from http://cmc-biostatistics.ac.in/Visali/Bootstrap
# The data has been saved as a csv file with the name "Bootstrap_data"
library(boot) # package to perform bootstrap
boot_paper<-read.csv("C:\\EY\\Visali_95%CI for Change outcome\\bootstrap_data.csv",sep=";",header=T)
#To use the package "Boot" it is mandatory to calculate the statistic on interest using a #function in R. The following function calculates the mean which will be called for each
#Bootstrap sample
meandiff <- function(y,i)
```

Table 2 – 95% CI using normal approximation, Bootstrap Bca using function and Bootstrap Percentile using the code.

Samples	95% CI using normal approximation	95% CI using BCa method from the boot function
$n = 15$	–159.94, 267.51	–114.81, 320.51
$n = 20$	–104.35, 242.61	–79.81, 255.70
$n = 25$	–42.92, 407.19	–1.70, 438.40
$n = 30$	29.19, 436.54	65.30, 455.40
$n = 35$	–45.58, 335.01	–25.80, 361.80
$n = 40$	–30.07, 312.58	–6.80, 336.50
$n = 50$	6.13, 294.47	22.00, 310.00
$n = 60$	56.59, 340.02	63.10, 341.90
$n = 75$	65.73, 299.44	76.00, 322.50

```

{
# function is written and this function # is named as "meandiff"
meand <- mean(y[i])      # calculating the mean
n<-length(y)             # sample size
c(meand,n)
}                          # calling the arguments passed in the function

```

```

# Bootstrapping with 10,000 replications
# The function mentioned above is called as shown in the
next line which provides means
# and n for 10,000 simulations
results <- boot(data=boot_paper$diff, statistic=meandiff,
R=10000)
boot.ci(results,type= c("perc", "bca")) # this produces
percentile and bca confidence intervals.

```

REFERENCES

1. Bland JM, Altman DG. Statistics notes: transforming data. *BMJ*. 1996;312:770.
2. Bland JM, Altman DG. Statistics notes: transformations, means, and confidence intervals. *BMJ*. 1996;312:1079.
3. Geyer Charles J. Stat 5102 Notes: *Nonparametric Tests and Confidence Intervals* [Internet]; 2003. Available from: <http://www.stat.umn.edu/geyer/old03/5102/notes/rank.pdf>.
4. Lehmann EL, D'Abrera HJM. *Nonparametrics: statistical Methods Based on Ranks*. New York: Springer; 2006.
5. DiCiccio TJ, Efrom Bradley. Bootstrap confidence intervals. *Stat Sci*. 1996;11:198–212.
6. The Comprehensive R Archive Network [Internet]. [cited 2012 Nov 23]. Available from: <http://cran.r-project.org/index.html>.
7. Efron B. Better bootstrap confidence intervals. *J Am Stat Assoc*. 1987;82:171–185.
8. Du Prel J-B, Hommel G, Röhrig B, Blettner M. Confidence interval or P-value? *Dtsch Arztebl Int*. 2009;106:335–339.
9. Hodges JL, Lehmann EL. Estimates of location based on rank tests. *Ann Math Stat*. 1963;34:598–611.
10. Diccio Thomas J, Romano Joseph P. A review of bootstrap confidence intervals. *J R Stat Soc*. 1988;50.
11. Carpenter J, Bithell J. Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians. *Stat Med*. 2000;19:1141–1164.
12. Felsenstein J. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*. 1985;39:783.
13. Wu CFJ. Jackknife, bootstrap and other resampling methods in regression analysis. *Ann Stat*. 1986;14:1261–1295.
14. Hinneburg A, Mannila H, Kaislaniemi S, Nevalainen T, Raumolin-Brunberg H. How to handle small samples: bootstrap and Bayesian methods in the analysis of linguistic change. *Lit Linguist Computing*. 2007;22:137–150.
15. Efron B, Tibshirani R. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Stat Sci*. 1986;1:54–75.