World Applied Sciences Journal 33 (5): 704-708, 2015
ISSN 1818-4952

DOI: 10.5829/idosi.wasj.2015.33.05.133

Box-Cox Transfromation and Bootstrapping Approach to One Sample T-Test

[1]*Wan Muhamad Amir W Ahmad,* [2]*Syerrina Binti Zakaria,*
[2]*Nor Azlida Aleng,* [2]*Nurfadhlina Abdul Halim and* [3]*Zalila Ali*

[1]School of Fisheries and Aquaculture Sciences,
University Malaysia Terengganu (UMT), Kuala Terengganu, Terengganu Malaysia
[1,2]School of Informatics and Applied Mathematics, University Malaysia Terengganu (UMT),
Kuala Terengganu, Terengganu Malaysia
[3]School of Mathematics Sciences, Universiti Sains Malaysia, 11800 Minden, Pulau Pinang, Malaysia

Abstract: One sample t-test is one of the most popular collections of statistical technique for analyzing data. Before we perform one sample T-Test the first thing that we should check is normality assumption. In this paper, we combine Box-Cox and bootstrapping idea in one algorithm. The purpose of Box-Cox is to ensure the data is normally distributed before the analysis. This combination is very useful for the modelling with an advanced analysis and perhaps can be an alternative method for modelling options in applied statistics scope. Through this combining method, we are capable to handle the case of non-normal data and small and limited sample size data by bootstrapping the original data set to generate new ones. In our case, the term "bootstrap" actually is referring to the use of the original data set to generate new ones. In this research paper, from a small and limited sample size data, we performed bootstrapping method in order to generate a new data set with a bigger sample size. After getting a new sample size, we then perform one sample T-Test using standard procedures and modified procedure. Results from both analyses will be compared with others to know the efficiency of the modified procedure. We also provided some example of application of the method discussed by using SAS language computer software.

Key words: Bootstrap・One Sample T-Test・Box-Cox Transformation

INTRODUCTION

The one-sample t-test is used to determine whether a sample comes from a population with a specific mean. This population mean is not always known, but is sometimes hypothesized. We do not know the population standard deviation, σ, but we know the sample standard deviation, s. There are some assumptions must be fulfill before performing the t test which are the sample must be random sample and the sample size is large ($n > 30$) or the population is approximately normal [1]. To test the hypothesis, we cannot use the z-score as our test statistic because we do not know, σ. Instead, we replace σ with the sample standard deviation, s and used the test statistic, $t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$ with $\bar{x}$ is the sample mean, μ is the population mean, s is the sample standard deviation and n is sample size [2].

However, the procedure have been discussed above is the normal procedure of one sample t test. The objective of this study is to discuss the improvement between the Box-Cox transformation and bootstrapping method for one sample t test with the normally procedure of one sample t test. Bootstrap method is a statistical technique that falls under the broad heading of resampling. This method is very useful and can be used various especially in the estimation of nearly any statistics [3]. This procedure involves a relatively simple procedure, but repeated so many times depending on the need of the researcher.

Bootstrap technique is heavily dependent upon computer calculation. Using the bootstrap method we are able to determine the estimating value of a parameter that presenting the whole of a population. Without using bootstrap method, the value of the parameter of a population is impossible to measure directly.

Corresponding Author: Wan Muhamad Amir W Ahmad, School of Fisheries and Aquaculture Sciences,
University Malaysia Terengganu (UMT), Kuala Terengganu, Terengganu Malaysia.

So, we use statistical sampling method and we sample a population, measure a statistic of the sample and then use these statistics to perfume one sample t test. Lastly, the results were interpreted and compare to the result for normally procedure of one sample t test. Normally distributed data is needed in order to use statistical analysis tools such as t-tests and analysis of variance. If data is not normally distributed, one of appropriate action should be taken is in transform the data to make data become normal. Data transformations are commonly-used tools to improve normality of a distribution and equalizing variance to meet assumptions and improve effect sizes, thus constituting important aspects of data cleaning and preparing for statistical analyses. There are as many potential types of data transformations. Some of the more commonly-discussed traditional transformations include: adding constants, square root, converting to logarithmic (e.g., base 10, natural log) scales, inverting and reflecting and applying trigonometric transformations such as sine wave transformations [4]. In this study, we are used the Box-Cox transformation. The form of Box-Cox transformation as below:

$$y(\lambda) = \begin{cases} \dfrac{y^{\lambda}-1}{\lambda}, & \text{if } \lambda \neq 0 \\ \log y, & \text{if } \lambda = 0 \end{cases}$$

Where, y is the observation data and λ is the model parameter. The optimal value of λ were determine and in this study, we used, $\lambda = 2$. We also provided some example of application of the method discussed by using SAS language computer software [4].

Algoritm Using Sas Language

Standard Procedure of One Sample Test:

```
Data time;
do I=1 to 20;
one=1; end;
input time;
datalines;
 43
 90
 84
 87
 116
 95
 86
 99
 93
 92
 121
 71
 66
 98
 79
 102
 60
 112
 105
 98
 ;
run;
ods rtf file='robdunc0.rtf' style=journal;

/* NORMALITY TEST */
Proc univariate normal plot data=time;
var time;

/* ONE SAMPLE T TEST*/
ods graphics on;
 proc ttest h0=80 plots(showh0) sides=u alpha=0.1;
 var time;
 run;
ods graphics off;
```

Modified Procedure of One Sample Test: One Sample T Test with Box-Cox Transformation and Bootstrapping Method

```
Data time;
do I=1 to 20;
one=1; end;
input time;
datalines;
 43
 90
 84
 87
 116
 95
 86
 99
 93
 92
 121
 71
 66
 98
 79
```

```
 102
 60
 112
 105
 98
 ;
 run;
ods rtf file='robdunc0.rtf' style=journal;
```

/* NORMALITY TEST */

```
Proc univariate normal plot data=time;
 var time;
```

/*IF THE DATA IS NOT NORMAL, SUGGESTION OF LAMBDA VALUE FOR TRANSFORMATION PROCEDURE WILL BE CONSIDER IN ORDER TO IMPROVE NORMALITY */

```
proc transreg data=time details;
 title2 'Defaults';
  model boxcox(time / lambda=-2 to 2 by 0.1) = identity(one);
 run;
```

/* BOX-COX SUGGEST (Y**(LAMBDA)-1)/LAMBDA WHEN THE LAMBDA VALUE IS NOT EQUAL TO ZERO*/

```
title1 'Transformed Variables';
data trans; set time;
 transformvalue= (time**(2)-1)/2;
proc print data = trans;
run;
```

/* ONE SAMPLE T TESTWITH BOOTSTRAP CASE RESAMPLING (REPLICATE =2) */

```
ods listing close;
proc surveyselect data = trans out = boot1 method = urs samprate =1 outhits rep=2;
run;
ods graphics on;
 proc ttest h0=80 plots(showh0) sides=u alpha=0.1;
 var time;
 run;
ods graphics off;
ods rtf close;
```

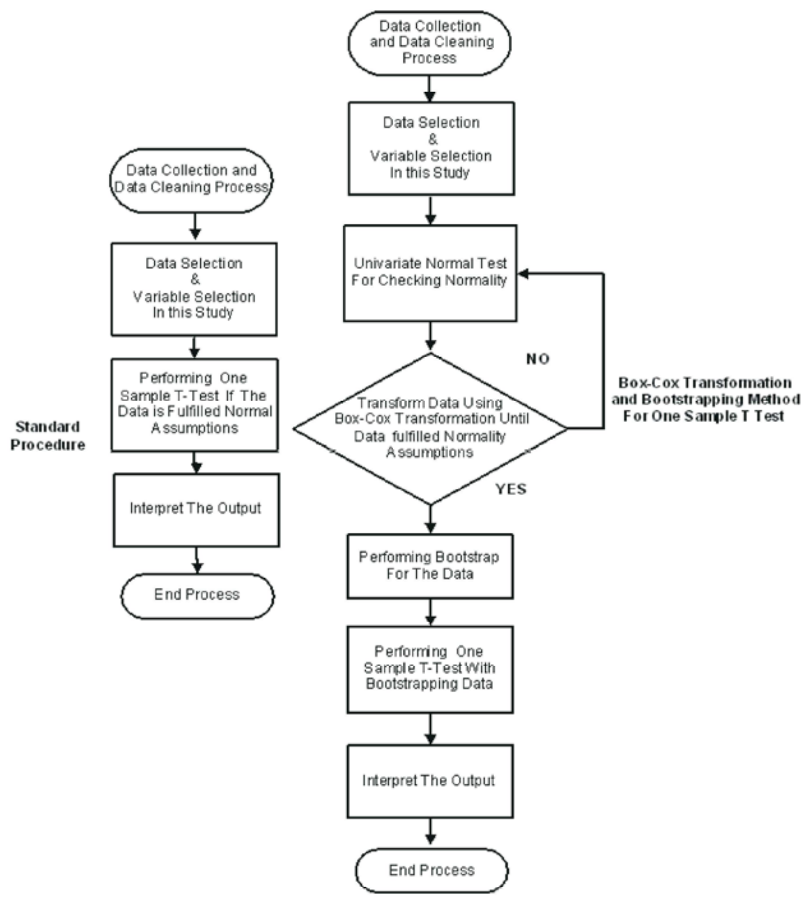

Fig. 1: Flow Chart of Normal Procedure and Alternative Analysis.

Figure 1 showed the flow chart of normal procedure and Box-Cox transformation and bootstrapping method procedure.

RESULTS

Standard Procedure:

Moments			
N	20	*Sum Weights*	20
Mean	89.85	*Sum Observations*	1797
Std Deviation	19.145633	*Variance*	366.555263
Skewness	-0.7081182	*Kurtosis*	0.66393627
Uncorrected SS	168425	*Corrected SS*	6964.55
Coeff Variation	21.3084396	*Std Error Mean*	4.28109369

Tests for Normality

Test	------ *Statistic* -----		-------- *p Value* -------	
Shapiro-Wilk	*W*	0.962943	*Pr < W*	0.6042
Kolmogorov-Smirnov	*D*	0.129973	*Pr > D*	>0.1500
Cramer-von Mises	*W-Sq*	0.054101	*Pr > W-Sq*	>0.2500
Anderson-Darling	*A-Sq*	0.310588	*Pr > A-Sq*	>0.2500

N	*Mean*	*Std Dev*	*Std Err*	*Minimum*	*Maximum*
20	89.8500	19.1456	4.2811	43.0000	121.0

Mean	*90% CL*	*Mean*	*Std Dev*	*90% CL*	*Std Dev*
89.8500	84.1659	Infty	19.1456	15.2002	26.2374

DF	*t Value*	*Pr > t*
19	2.30	0.0164

One Sample T Test with Box-Cox Transformation and Bootstrapping Method

The TRANSREG Procedure		
Box-Cox Lambda	*Transformation R-Square*	*Information for time Log Like*
-2.0	0.00	-71.4259
-1.9	0.00	-70.7456
-1.8	0.00	-70.0825
-1.7	0.00	-69.4368
-1.6	0.00	-68.8089
-1.5	0.00	-68.1990
-1.4	0.00	-67.6073
-1.3	0.00	-67.0341
-1.2	0.00	-66.4795
-1.1	0.00	-65.9438
-1.0	0.00	-65.4269
-0.9	0.00	-64.9292
-0.8	0.00	-64.4505
-0.7	0.00	-63.9911
-0.6	0.00	-63.5509
-0.5	0.00	-63.1299
-0.4	0.00	-62.7281
-0.3	0.00	-62.3455
-0.2	0.00	-61.9819
-0.1	0.00	-61.6372
0.0	0.00	-61.3113
0.1	0.00	-61.0041

Continued

The TRANSREG Procedure		
Box-Cox Lambda	*Transformation R-Square*	*Information for time Log Like*
0.2	0.00	-60.7153
0.3	0.00	-60.4448
0.4	0.00	-60.1924*
0.5	0.00	-59.9577*
0.6	0.00	-59.7405*
0.7	0.00	-59.5406*
0.8	0.00	-59.3577*
0.9	0.00	-59.1914*
1.0+	0.00	-59.0415*
1.1	0.00	-58.9076*
1.2	0.00	-58.7895*
1.3	0.00	-58.6867*
1.4	0.00	-58.5990*
1.5	0.00	-58.5260*
1.6	0.00	-58.4674*
1.7	0.00	-58.4229*
1.8	0.00	-58.3920*
1.9	0.00	-58.3746*
2.0	0.00	-58.3701<

< - Best Lambda
** - 95% Confidence Interval*
+ - Convenient Lambda

The TRANSREG Procedure

Model Statement Specification Details				
Type	*DF*	*Variable*	*Description*	*Value*
Dep	1	BoxCox(time)	Lambda Used	2
			Lambda	2
			Log Likelihood	-58.3701
			Conv. Lambda	1
			Conv. Lambda LL	-59.0415
			CI Limit	-60.2909
			Alpha	0.05
Ind	0	Identity(one)	Options	All Zero

Transformed Variables

Obs	*I*	*one*	*time*	*transformvalue*
1	21	1	43	924.0
2	21	1	90	4049.5
3	21	1	84	3527.5
4	21	1	87	3784.0
5	21	1	116	6727.5
6	21	1	95	4512.0
7	21	1	86	3697.5
8	21	1	99	4900.0
9	21	1	93	4324.0
10	21	1	92	4231.5
11	21	1	121	7320.0
12	21	1	71	2520.0
13	21	1	66	2177.5
14	21	1	98	4801.5
15	21	1	79	3120.0
16	21	1	102	5201.5
17	21	1	60	1799.5
18	21	1	112	6271.5
19	21	1	105	5512.0
20	21	1	98	4801.5

The SURVEYSELECT Procedure

Input Data Set	TRANS
Random Number Seed	311622001
Sampling Rate	1
Sample Size	20
Expected Number of Hits	1
Sampling Weight	1
Number of Replicates	2
Total Sample Size	40
Output Data Set	BOOT1

Transformed Variables

The TTEST Procedure

N	*Mean*	*Std Dev*	*Std Err*	*Minimum*	*Maximum*
40	89.1250	20.5191	3.2444	43.0000	121.0

Mean	*90% CL*	*Mean*	*Std Dev*	*90% CL*	*Std Dev*
89.1250	84.8955	Infty	20.5191	17.3463	25.2792

DF	*t Value*	*Pr > t*
39	2.81	0.0038

Based on normal procedure, the one sample t-test statistic is 2.30 and the *p*-value from this statistic is 0.0164 and that is less than 0.05 significant level. Such a *p*-value indicates that the average of the sampled population is statistically significantly different from 80. Compared to the alternative procedure, the t-test statistic is 2.81 with 0.0038 *p*-value. It was indicate that the mean of the data is statistically significantly different from 80. It showed that the result from these two procedures is similar but the *p*-value for the alternative procedure is smaller. From the results, we can say that about 76.8% of the alternative procedure are much better at giving a good result compared to the standard method. Below is the comparison of the gained results:

Standard Proceduret			*Alternative Procedure*		
DF	*t Value*	*Pr > t*	*DF*	*t Value*	*Pr > t*
19	2.30	0.0164	39	2.81	0.0038

DISCUSSION

This paper explained the comparison between standard procedure and Box-Cox transformation and Bootstrapping method for one sample t-test. The bootstrap method offered a preliminary general idea of the process that involving inadequate sample size and straightly solve the problem by bootstrapping the observations thus exceeding the minimum requirements of the sample size. In this study, two different methods have been used: (i) normal procedure for one sample t-test and (ii) Box-Cox transformation and Bootstrapping method for one sample t-test. The first case study examined data using normal procedure and the second case study analyses using transformation the data and bootstrapping method of enlarging the sample size. From the both methods, it shows that the mean of the data is statistically significant from 80. But the most different of, the p-value for the second method is smaller than the p-value of normal method. Through this combining method, we are capable to handle the case of non-normal data and small and limited sample size data by bootstrapping the original data set to generate new ones.We can conclude that the second method is more compatible with a minuscule sample size and able to show the smoother normal distribution.

REFERENCES

1. Bluman, A.G., 2014. Elementary Statistics: A Step by Step Approach, Ninth Edition, McGraw-Hill International Edition
2. Navidi, W. and B. Monk, 2013. Elementary Statistics, McGraw-Hill International Edition.
3. Cassel, D.L., 2010. Bootstrap Mania: Re sampling the SAS. SAS Global Forum 2010: Statistics and Data Analysis. Paper 268-2010: pp: 1-11.
4. Osborne, J.W., 2010. Improving your data transformations?: Applying the Box-Cox transformation. Practical Assessment, Research & Evaluation, 15(12): 1-9.
5. O'Rourke, N., E.J. Stepanski and L. Hatcher, 2005. A Step-by-step approach to using SAS for univariate & multivariate statistics, New Jersey John Wiley & Sons 2nd Edition.