Transformations for Left Skewed Data

Lakhana Watthanacheewakul

Abstract—**The normality is an important assumption in the statistical methods. Thus, we should investigate the distribution of data before analyzing data. If the original data do not correspond with normality, they will be transformed to normality. A simple mathematical function can transform only some non-normal data sets to normality such as square root, logarithm, and inverse. Hence a family of transformation is used to transform non-normal data such as Box-Cox transformation, Manly transformation, and Yeo-Johnson transformation. These transformations are commonly used via statistical computing program. Some distributions have both left and right skewed such as Weibull distribution, Beta distribution, and so on. There are different methods for transforming left skewed data to normality. The objective of this paper is to compare eight different methods used for transforming left skewed Weibull data and left skewed Beta data to normality: reflect then logarithm with base 10 transformation, reflect then square root transformation, Box-Cox transformation, reflect then Box-Cox transformation, Manly transformation, reflect then Manly transformation, Yeo-Johnson transformation, and reflect Yeo-Johnson transformation in sense of reducing skewness, normality, and maintaining dispersion. R programming language is used to generate left skewed Weibull data and left skewed Beta data including data processing. In conclusion, left skewed Weibull data were reflected then transformed by Yeo-Johnson transformation and left skewed Weibull data were reflected then Manly transformation had a good performance in sense of normality, reducing skewness and maintaining dispersion in every situation. For left skewed Beta data, they were reflected then transformed by Manly transformation had a good performance in sense of normality, reducing skewness and dispersion. Although, some transformations can transform both left skewed Weibull data and left skewed Beta data to normality but the level of skewness of transformed data was not symmetry and the dispersion of the transformed data was different from the original data over than 20 percent.**

Index Terms—left skewed data, Box-Cox transformation, Manly transformation, Yeo-Johnson transformation

Manuscript received February 6, 2020; revised March 26, 2020.
L. Watthanacheewakul is with the Faculty of Science, Maejo University, Chiang Mai, Thailand (phone: 66-53-873-881; fax: 66-53-873-827; e-mail: lakhanaw@yahoo.com).

I. Introduction

DATA analysis is a necessary process in research methodology, especially in quantitative research. The normality is an essential assumption in the most statistical methods. If the mean, median and mode of data are all the same, the distribution will be symmetric or normally. If they are all different, the distribution will be skewed. Thus, we should investigate the distribution of data before analyzing data. The coefficient of skewness is one of many ways to investigate the distribution of data. If the value of it is positive, the data have right skewed distribution. If the value of it is negative, the data have left skewed distribution. Some non- normal distributions can be either left skewed or right skewed such as Weibull distribution, Beta distribution, and so on. Pyzdek [1] illustrated how the non-normal quality characteristic data would significantly impact the data analysis result and the conclusion. Tukey [2] suggested that there are two methods; transform the data to fit the assumptions or develop some new robust methods of analysis when data do not match the assumptions of a traditional method of analysis. Wuensch [3] suggested that the positive skewness is reduced by the simple mathematical functions such as logarithm, square root, and square. If the skewness is negative, reflection technique will require prior to transformation. Reflection means each observation is subtracted from a constant that is higher than the highest observation. However, they cannot transform some non-normal data set to normality. Cox [4] indicated that we can use higher powers to reduce left skewness. Hence A family of transformations studied over a long period of time, e.g. Box and Cox [5], Manly [6], and Yeo and Johnson [7] can transform them to normality. The normality is considered with Lilliefors test and the skewness is measured by the coefficient of skewness (C.S.). Moreover, the dispersion of the transformed data and the original data should have closed, it is measured by the coefficient of variation (C.V.). In this paper, we compared eight methods for transforming the left skewed data; reflect then logarithm with base 10 transformation (RL), reflect then square root transformation (RR), Box-Cox transformation (BC), reflect then Box-Cox transformation (RBC), Manly transformation (M), reflect then Manly transformation (RM), Yeo-Johnson transformation (YJ), and reflect then Yeo-Johnson transformation (RYJ) in sense of reducing skewness, normality, and maintaining dispersion. R programming language [8] is used in statistical computing.

A. *Traditional Transformation*

Baker [9] divided the skewness of distribution into moderate, high and extreme and introduced the traditional transformation as Table I

ISBN: 978-988-14049-2-3
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

TABLE I
TRADITIONAL TRANSFORMATION FOR LEFT SKEWED DISTRIBUTION AND RIGHT SKEWED DISTRIBUTION

Skewness	Right skewed distribution	Left skewed distribution
Moderate	Square root transformation	Reflect then square root transformation
High	Natural logarithm transformation	Reflect then Natural Logarithm transformation
Higher	Logarithm base 10 transformation	Reflect then Logarithm base 10 transformation
Extreme	Inverse transformation	Reflect then Inverse transformation

Source: Transforming Skewed Data (Baker, 2017)

B. A Family of Transformations

Let X be a random variable distributed as non-normal, Y the transformed variable of X, x the value of X, and λ a transformation parameter.

Box and Cox [5] proposed a family of transformations in this form

$$Y = \begin{cases} \dfrac{X^{\lambda}-1}{\lambda}, & \lambda \neq 0 \\ \ln X, & \lambda = 0 \end{cases} \quad \text{for} \quad x > 0. \tag{1}$$

It has known as Box-Cox transformation.

Manly [6] proposed a family of transformations in this form

$$Y = \begin{cases} \dfrac{\exp(\lambda X)-1}{\lambda}, & \lambda \neq 0 \\ X, & \lambda = 0. \end{cases} \tag{2}$$

It has known as exponential transformation. The value of X can be either positive and negative.

Yeo and Johnson [7] proposed a family of transformations in this form

$$Y = \begin{cases} \dfrac{[X+1]^{\lambda}-1}{\lambda} & , x \geq 0,\ \lambda \neq 0 \\ \ln[X+1] & , x \geq 0,\ \lambda = 0 \\ -\left[(-X+1)^{2-\lambda}-1\right] \big/ 2-\lambda & , x < 0,\ \lambda \neq 2 \\ -\ln(-X+1) & , x < 0,\ \lambda = 2 \end{cases} \tag{3}$$

It has known as Yeo-Johnson transformation.

II. LEFT SKEWED DATA

For left skewed data, the tail of distribution is on the left and the coefficient of skewness is negative. The distribution is moderately skewed if the coefficient of skewness is between -0.5 and -1, the distribution is highly skewed if the coefficient of skewness is less than -1. (Jain [10]).

The central tendency measures in a left skewed data may be illustrated as follows: Mode > Median > Mean. For example, if a final exam is easy, then most of the students will have high scores and only few students will have low scores. Corporate Finance Institute [11] indicated that many trading strategies employed by traders are based on left skewed data (small gains and large losses).

Weibull distribution and Beta distribution are involved left skewed data. The Weibull distribution is broadly applied in reliability and lifetime data. It depends on the values of the shape parameter, β, and the scale parameter, α. The probability density function of a two parameter Weibull random variable X is

$$f(x) = \frac{\beta}{\alpha}\left(\frac{x}{\alpha}\right)^{\beta-1} e^{-\left(\frac{x}{\alpha}\right)^{\beta}},\ x \geq 0;\ \alpha, \beta > 0. \tag{4}$$

The reliability professional [12] suggested that if the shape parameter is greater than 3.7, it will be left skewed distribution as Fig. 1.

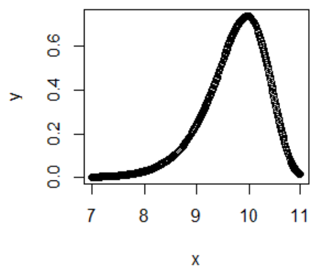

Fig. 1. Graph of Weibull distributions when shape parameter is 20 and scale parameter is 10

The Beta distribution is distribution on the interval [0,1]. The probability density function of a two parameter Beta random variable X is

$$f(x) = \frac{\Gamma(\gamma+\beta)}{\Gamma(\gamma)\Gamma(\beta)} x^{\gamma-1}(1-x)^{\beta-1},\ 0 < x < 1;\ \gamma, \beta > 0 \tag{5}$$

where $\Gamma(.)$ is the Gamma function, and both α and β are the shape parameters. The skewness can be either left or right. For example, when γ =5 and β =2, the Beta distribution is left skewed as Fig. 2 (Ma [13]).

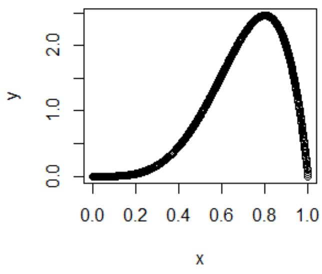

Fig.2. Graph of Beta distributions when shape parameters are 5 and 2

III. EXAMPLE OF REAL DATA

From Bank of Thailand, data of the return on investment of non-bank fromQ1/2015 to Q2/2019 were as Table II.

ISBN: 978-988-14049-2-3
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

TABLE II
THE RETURN ON INVESTMENT OF NON-BANK
FROM Q1/2015 TO Q2/2019

Year	ROI (%)	Year	ROI (%)
Q1/2015	39.9	Q2/2017	40.0
Q2/2015	39.6	Q3/2017	38.5
Q3/2015	39.8	Q4/2017	38.7
Q4/2015	40.1	Q1/2018	37.6
Q1/2016	39.2	Q2/2018	38.7
Q2/2016	39.5	Q3/2018	38.8
Q3/2016	39.7	Q4/2018	39.6
Q4/2016	40.4	Q1/2019	38.6
Q1/2017	39.7	Q2/2019	39.5

Source: Bank of Thailand [14]

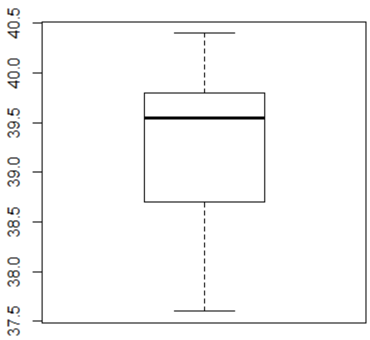

Fig.3. Boxplot of Data in Table II

From Fig. 3, the data had left skewed distribution. The data were transformed to normality by eight methods of transformation. The results from R programming language were as Table III.

TABLE III
TRANSFORMATION PARAMETER, COEFFICIENT OF SKEWNESS, COEFFICIENT OF VARIATION, AND P-VALUE OF LILLIEFORS TEST
IN EACH TRANSFORMATION

Transformation	Transformation parameter	C.S.	C.V.	P-value Lilliefors
None	-	-0.7360	0.0180	0.0397
RL	-	0.0407	0.4979	0.3447
RR	-	0.3965	0.1691	0.1249
BC	2	-0.6973	**0.0358**	0.0446
M	0.4713	-0.0820	0.3000	0.2999
YJ	5	-0.5876	0.0855	0.0624
RBC	-0.1	-0.0333	0.4850	**0.4111**
RM	-0.4713	0.0820	0.1961	0.2999
RYJ	-0.5636	**0.0237**	0.1464	0.3609

From Table III, the original data had left skewed distribution and non-normality at significant level 0.05. The data were reflected then transformed by Yeo-Johnson transformation are the most efficient in sense of reducing skewness. Box-Cox transformation can maintain dispersion but it cannot transform the data to normality at significant level 0.05. Although, the data were reflected then transformed by Box-Cox transformation have performance in sense of reducing skewness and normality but dispersion of transformed data was far from its of original data over than 45%. In sense of reducing skewness, normality and maintaining dispersion, the data were reflected then transformed by Yeo-Johnson transformation, the data were reflected then transformed by Manly transformation and the data were reflected then transformed by square root are interested, respectively.

IV. SIMULATION STUDY

The scope of this work was as follows:

1) Weibull data and Beta data were left skewed in moderate and high level.
2) Lilliefors test was used for testing the normality.
3) There were eight methods for transforming the left skewed data to normality; reflect then logarithm base 10 transformation (RL), reflect then square root transformation (RR), Box-Cox transformation (BC), reflect then Box-Cox transformation (RBC), Manly transformation (M), reflect then Manly transformation (RM), Yeo - Johnson transformation (YJ), and reflect then Yeo - Johnson transformation (RYJ).
4) The transformed data were considered in sense of reducing skewness, normality, and dispersion.
5) There were three levels of sample size (n): the small sample size (n=10, 20), the medium sample size (n=30, 50), and the large sample size (n=100, 150).
6) The significant level was set at 0.05.

The process of this work was as follows:

1) For Weibull data, the Weibull populations of size N=5,000 were generated for moderately skewed and highly skewed (the scale parameter=20 and the shape parameter=5), then 1,000 random samples for each sample size, are drawn. Each set of the sample data was transformed to normality by eight transformation methods. The results of transformed data in form of the average of the coefficient of skewness, the average of the coefficient of variation, and the average of the p-value of Lilliefors test are shown in Table IV.
2) For Beta data, the Beta populations of size N=5,000 were generated for moderately skewed and highly skewed (the first shape parameter=8 and the second shape parameter=2), then 1,000 random samples for each sample size, are drawn. Each set of the sample data was transformed to normality by eight transformation methods. The results of transformed data in form of the average of the coefficient of skewness, the average of the coefficient of variation, and the average of the p-value of Lilliefors test are shown in Table V.

ISBN: 978-988-14049-2-3
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

TABLE IV
AVERAGE OF COEFFICIENT OF SKEWNESS, AVERAGE OF COEFFICIENT OF VARIATION, AND AVERAGE OF P-VALUE OF LILLIEFORS TEST BY TRANSFORMATIONS FOR WEIBULL DATA

n	Level of skewness	Transformation	C.S.	C.V.	P-value Lilliefors
10	moderate	None	-0.7746	0.0699	0.0204
		RL	0.4437	0.7575	0.0604
		RR	0.6140	0.1176	0.0349
		BC	-0.6750	0.1416	0.0279
		M	-0.1849	0.4011	0.1713
		YJ	-0.1032	0.2629	0.0595
		RBC	0.1357	0.6536	0.2147
		RM	**0.1849**	**0.0714**	**0.1713**
		RYJ	**0.1508**	**0.0404**	**0.2058**
	high	None	-1.4708	0.0858	0.0095
		RL	1.1261	0.9035	0.0402
		RR	1.3112	0.1365	0.0188
		BC	-1.3605	0.1664	0.0148
		M	-0.3298	0.5136	0.4488
		YJ	-1.0766	0.2817	0.0479
		RBC	0.3652	0.6839	0.3316
		RM	**0.3298**	**0.0379**	**0.4488**
		RYJ	**0.3249**	**0.0166**	**0.4053**
20	moderate	None	-0.7136	0.0580	0.0414
		RL	0.3105	0.5235	0.1182
		RR	0.5140	0.0950	0.0688
		BC	-0.5884	0.1181	0.0547
		M	-0.0648	0.2827	0.2047
		YJ	-0.3099	0.2225	0.1092
		RBC	0.0202	0.4683	0.2387
		RM	**0.0648**	**0.0739**	**0.2047**
		RYJ	**0.0375**	**0.0475**	**0.2255**
	high	None	-1.9053	0.0730	0.0236
		RL	1.0889	0.7002	0.1884
		RR	1.5009	0.1165	0.0748
		BC	-1.6074	0.1406	0.0564
		M	-0.0811	0.3878	0.4324
		YJ	-0.9563	0.2419	0.2763
		RBC	0.0307	0.5497	0.3630
		RM	**0.0811**	**0.0546**	**0.4324**
		RYJ	**0.0358**	**0.0267**	**0.3860**
30	moderate	None	-0.9015	0.0609	0.0237
		RL	0.3912	0.6177	0.0930
		RR	0.6404	0.1025	0.0475
		BC	-0.7287	0.1233	0.0359
		M	-0.0862	0.3155	0.1437
		YJ	-0.3702	0.2310	0.0998
		RBC	0.0454	0.5478	0.1738
		RM	**0.0862**	**0.0743**	**0.1437**
		RYJ	**0.0650**	**0.0458**	**0.1699**
	high	None	-1.4482	0.0813	0.0090
		RL	0.8268	0.6861	0.0745
		RR	1.1353	0.1267	0.0261
		BC	-1.2177	0.1592	0.0179
		M	-0.1568	0.4176	0.5406
		YJ	-0.7387	0.1879	0.1755
		RBC	0.0643	0.5416	0.5153
		RM	**0.1568**	**0.0659**	**0.3160**
		RYJ	**0.0999**	**0.0328**	**0.5488**

TABLE IV (CONTINUE)

n	Level of skewness	Transformation	C.S.	C.V.	P-value Lilliefors
50	moderate	None	-0.8024	0.0574	0.0174
		RL	0.3449	0.5130	0.1675
		RR	0.5754	0.0940	0.0587
		BC	-0.6576	0.1167	0.0392
		M	-0.0467	0.2821	0.4910
		YJ	-0.3328	0.2201	0.1848
		RBC	-0.0038	0.4564	0.5306
		RM	**0.0467**	**0.0727**	**0.4910**
		RYJ	**0.0240**	**0.0474**	**0.5066**
	high	None	-1.3415	0.0588	0.0261
		RL	0.5279	0.3980	0.2795
		RR	0.9276	0.0876	0.0980
		BC	-1.0343	0.1175	0.0740
		M	0.0637	0.2742	0.3628
		YJ	-0.4097	0.2162	0.3666
		RBC	-0.0602	0.3386	0.2831
		RM	**-0.0637**	**0.0612**	**0.3628**
		RYJ	**-0.0593**	**0.0396**	**0.3164**
100	moderate	None	-0.8341	0.0568	0.0224
		RL	0.4191	0.5036	0.3484
		RR	0.6232	0.0932	0.1003
		BC	-0.6935	0.1154	0.0558
		M	-0.0832	0.3086	0.6707
		YJ	-0.3986	0.2177	0.3340
		RBC	0.0250	0.4260	0.7468
		RM	**0.0832**	**0.0628**	**0.6707**
		RYJ	**0.0521**	**0.0367**	**0.7185**
	high	None	-1.6066	0.0630	0.0119
		RL	0.6771	0.4634	0.1966
		RR	1.1031	0.0966	0.0458
		BC	-1.1897	0.1246	0.0294
		M	-0.0195	0.3149	0.5788
		YJ	-0.5159	0.2271	0.2861
		RBC	-0.0116	0.3760	0.6112
		RM	**0.0195**	**0.0608**	**0.5788**
		RYJ	**-0.0033**	**0.0346**	**0.5933**
150	moderate	None	-0.7390	0.0661	0.0117
		RL	0.2060	0.4076	0.4717
		RR	0.4730	0.0973	0.1154
		BC	-0.5468	0.1340	0.0628
		M	-0.0280	0.2469	0.7170
		YJ	-0.1544	0.2471	0.5231
		RBC	0.0122	0.3777	0.6799
		RM	**0.0280**	**0.0917**	**0.7170**
		RYJ	**0.0089**	**0.0663**	**0.7085**
	high	None	-1.5591	0.0672	0.0034
		RL	0.5975	0.4706	0.3476
		RR	1.0236	0.1016	0.0601
		BC	-1.0959	0.1327	0.0301
		M	-0.0363	0.3107	0.8475
		YJ	-0.4213	0.2420	0.4279
		RBC	-0.0190	0.3871	0.7700
		RM	**0.0363**	**0.0705**	**0.8475**
		RYJ	**0.0111**	**0.0413**	**0.8115**

TABLE V
AVERAGE OF COEFFICIENT OF SKEWNESS, AVERAGE OF COEFFICIENT OF VARIATION, AND AVERAGE OF P-VALUE OF LILLIEFORS TEST BY TRANSFORMATIONS FOR BETA DATA

n	Level of skewness	Transformation	C.S.	C.V.	P-value Lilliefors
10	moderate	None	-0.6940	0.1408	0.0228
		RL	0.6492	0.9398	0.0300
		RR	0.6721	0.0508	0.0261
		BC	-0.6290	-0.5889	0.0331
		M	-0.4592	0.3975	0.0862
		YJ	-0.5742	0.2997	0.0459
		RBC	0.5466	0.8652	0.0550
		RM	**0.4592**	**0.0049**	**0.0862**
		RYJ	0.5370	0.0061	0.0577
	high	None	-1.3151	0.1982	0.0247
		RL	1.1478	0.9144	0.0628
		RR	1.2338	0.0659	0.0397
		BC	-1.0082	-0.5308	0.1188
		M	-0.4167	0.4603	0.6148
		YJ	-0.8061	0.3730	0.2515
		RBC	0.7674	0.7788	0.2750
		RM	**0.4167**	**0.0050**	**0.6148**
		RYJ	0.7158	0.0067	0.3244
20	moderate	None	-0.6933	0.1636	0.0393
		RL	0.5727	0.7577	0.0754
		RR	0.6333	0.0559	0.0540
		BC	-0.5065	-0.6059	0.1052
		M	-0.1935	0.4123	0.3431
		YJ	-0.3708	0.3407	0.2184
		RBC	0.3256	0.6697	0.2745
		RM	**0.1935**	**0.0075**	**0.3431**
		RYJ	**0.2963**	**0.0062**	**0.3005**
	high	None	-1.2526	0.1713	0.0191
		RL	1.0377	0.8100	0.0562
		RR	1.1456	0.0575	0.0331
		BC	-0.8867	-0.5796	0.1070
		M	-0.2395	0.4260	0.7471
		YJ	-0.6556	0.3289	0.2626
		RBC	0.6034	0.6950	0.3136
		RM	**0.2395**	**0.0046**	**0.7471**
		RYJ	0.5469	0.0060	0.3737
30	moderate	None	-0.8072	0.1557	0.0145
		RL	0.6486	0.6844	0.0371
		RR	0.7282	0.0523	0.0232
		BC	-0.5543	-0.5363	0.0580
		M	-0.1443	0.3923	0.4468
		YJ	-0.3767	0.3216	0.1546
		RBC	0.3236	0.5939	0.2112
		RM	**0.1443**	**0.0069**	**0.4468**
		RYJ	**0.2836**	**0.0057**	**0.2503**
	high	None	-1.4803	0.1451	0.0070
		RL	1.2207	0.6894	0.0234
		RR	1.3514	0.0486	0.0129
		BC	-1.0285	-0.4716	0.0509
		M	-0.2739	0.3675	0.4875
		YJ	-0.7495	0.2812	0.1420
		RBC	0.6994	0.5796	0.1635
		RM	**0.2739**	**0.0039**	**0.4875**
		RYJ	0.6304	0.0051	0.2013

TABLE V (CONTINUE)

n	Level of skewness	Transformation	C.S.	C.V.	P-value Lilliefors
50	moderate	None	-0.8118	0.1421	0.0062
		RL	0.6304	0.6679	0.0258
		RR	0.7203	0.0487	0.0128
		BC	-0.5209	-0.5571	0.0527
		M	-0.1265	0.3641	0.4592
		YJ	-0.3370	0.2978	0.1962
		RBC	0.2873	0.5842	0.2613
		RM	**0.1265**	**0.0070**	**0.4592**
		RYJ	**0.2454**	**0.0054**	**0.3163**
	high	None	-1.0805	0.1511	0.0165
		RL	0.7935	0.6274	0.0605
		RR	0.9337	0.0501	0.0315
		BC	-0.6015	-0.5061	0.1152
		M	-0.0784	0.3611	0.4392
		YJ	-0.3408	0.3075	0.3055
		RBC	0.2890	0.5345	0.3766
		RM	**0.0784**	**0.0078**	**0.4392**
		RYJ	**0.2261**	**0.0053**	**0.3917**
100	moderate	None	-0.9322	0.1506	0.0153
		RL	0.6568	0.5833	0.0939
		RR	0.7911	0.0493	0.0387
		BC	-0.4721	-0.4980	0.1514
		M	-0.0693	0.3397	0.7111
		YJ	-0.2275	0.3099	0.3777
		RBC	0.1835	0.4974	0.5127
		RM	**0.0693**	**0.0115**	**0.7111**
		RYJ	**0.1631**	**0.0067**	**0.5346**
	high	None	-1.2658	0.1574	0.0244
		RL	0.9105	0.6288	0.1633
		RR	1.0813	0.0520	0.0778
		BC	-0.6558	-0.5439	0.3749
		M	-0.0925	0.3718	0.5411
		YJ	-0.3750	0.3137	0.5632
		RBC	0.3414	0.5285	0.6283
		RM	**0.0925**	**0.0074**	**0.5411**
		RYJ	**0.2682**	**0.0054**	**0.5988**
150	moderate	None	-0.7945	0.1492	0.0097
		RL	0.5495	0.5758	0.0601
		RR	0.6685	0.0492	0.0272
		BC	-0.3922	-0.5083	0.1306
		M	-0.0817	0.3234	0.1681
		YJ	-0.1733	0.3135	0.2518
		RBC	0.1319	0.4920	0.3086
		RM	**0.0817**	**0.0130**	**0.1681**
		RYJ	**0.0990**	**0.0059**	**0.2305**
	high	None	-1.1725	0.1525	0.0120
		RL	0.8295	0.5750	0.1046
		RR	0.9959	0.0499	0.0360
		BC	-0.5876	-0.5112	0.2627
		M	-0.0595	0.3486	0.8264
		YJ	-0.2940	0.3082	0.7069
		RBC	0.2553	0.4806	0.8069
		RM	**0.0595**	**0.0086**	**0.8264**
		RYJ	**0.2380**	**0.0052**	**0.7458**

ISBN: 978-988-14049-2-3
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

From Table IV, for reducing skewness, M RBC RM and RYJ can reduce skewness to symmetry (the coefficient of skewness is between -0.5 and 0.5) in every situation. For maintaining dispersion, RR BC RM and RYJ can maintain dispersion. The coefficient of variation of transformed data was not over than 10 % from its of original data in every situation. For normality, M RBC RM and RYJ can transform left skewed Weibull data to normality in every situation at significant level 0.05. Sample size had affected the p-value of Lilliefors test.

From Table V, for reducing skewness, M and RM can reduce skewness to symmetry in every situation. For maintaining dispersion, RR RM YJ and RYJ can maintain dispersion. The coefficient of variation of transformed data was not over than 20% from its of original data in every situation. Although the coefficient of variation of transformed data by RR was different from its of original data about 10-14 % in each situation but RR cannot transform the left skewed Beta data to normality and RR cannot reduce the skewness. For normality, M RBC RM and RYJ can transform left skewed Beta data to normality in every situation at significant level 0.05. Sample size and level of skewness of original data had affected the coefficient of skewness, the coefficient of variation, and the p-value of Lilliefors test.

V. Conclusion

For the left skewed Weibull data, the data reflected then transformed by Manly transformation and the data reflected then transformed by Yeo-Johnson transformation had a good performance in sense of normality, symmetry and dispersion in every situation.

For the left skewed Beta data, the data reflected then transformed by Manly transformation had a good performance in sense of normality, symmetry and dispersion in every situation.

References

[1] T. Pyzdek, "Normal Distributions Aren't [All That Normal]," *Quality Engineering.* vol. 7, no. 4, pp. 769-777, Jun. 1995.
[2] J. W. Tukey, "On the comparative anatomy of transformations," *Annals of Mathematical Statistics*, vol. 28, no. 3, pp.602-632, Sep. 1957.
[3] K. L. Wuensch. (2017, November).Transforming to Reduce Negative Skewness.Available:http://core.ecu.edu/psyc/wuenschk/StatHelp/NegSkew.pdf
[4] N. J. Cox. (2007, July).Transformations: an introduction. Available: http://fmwww.bc.edu/repec/bocode/t/transint.html
[5] G. E. P. Box and D. R. Cox, "An analysis of transformations (with discussion)," *Journal of the Royal Statistical Society, Ser.B*. vol. 26, no. 2, pp.211-252, Apr. 1964.
[6] B. F. J. Manly, "Exponential Data Transformations," *Statistician*. vol. 25, no. 1, pp.37-42, Mar. 1976.
[7] I. Yeo and N. R. Johnson, "A new family of power transformations to improve normality or symmetry," *Biometrika*, vol. 87, no. 2, pp. 954-959, Dec. 2000.
[8] R Core Team. (2019). "R: A language and environment for statistical computing," *R Foundation for Statistical Computing*, Vienna, Austria. Available: https://www.R-project.org/
[9] S. S. Baker. (2017) "Transforming Skewed Data," Available: https://www.anatomisebiostats.com/biostatistics-blog/transforming-skewed-data
[10] D. Jain. (2018). "Skew and Kurtosis: 2 Important Statistics terms you need to know in Data Science," Available: https://codeburst.io/2-important-statistics-terms-you-need-to-know-in-data-science-skewness-and-kurtosis-388fef94eeaa
[11] Corporate Finance Institute, "Negatively Skewed Distribution," Available:https://corporatefinanceinstitute.com/resources/knowledge/other/negatively-skewed-distribution/
[12] The Reliability Professional, "Weibull Distribution Characteristics," Available:http://www.reliawiki.com/index.php/Weibull_Distribution_Characteristics
[13] D. Ma. (2015). " The skewness of a probability distribution," Available: https://probabilityandstats.wordpress.com/tag/beta-distribution/
[14] Bank of Thailand. (2020). "Financial Soundness Indicators" Available:https://www.bot.or.th/App/BTWS_STAT/statistics/ReportPage.aspx?reportID=858&language=th

ISBN: 978-988-14049-2-3
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)