International Journal of Mechanical, Industrial and Aerospace Sciences
ISSN: 2517-9950
Vol:6, No:12, 2012

Comparative Study of Transformed and Concealed Data in Experimental Designs and Analyses

K. Chinda, and P. Luangpaiboon

Abstract**—**This paper presents the comparative study of coded data methods for finding the benefit of concealing the natural data which is the mercantile secret. Influential parameters of the number of replicates (*rep*), treatment effects (τ) and standard deviation (σ) against the efficiency of each transformation method are investigated. The experimental data are generated via computer simulations under the specified condition of the process with the completely randomized design (CRD). Three ways of data transformation consist of Box-Cox, arcsine and logit methods. The difference values of F statistic between coded data and natural data (F_c-F_n) and hypothesis testing results were determined. The experimental results indicate that the Box-Cox results are significantly different from natural data in cases of smaller levels of replicates and seem to be improper when the parameter of minus lambda has been assigned. On the other hand, arcsine and logit transformations are more robust and obviously, provide more precise numerical results. In addition, the alternate ways to select the lambda in the power transformation are also offered to achieve much more appropriate outcomes.

Keywords**—**Experimental Designs, Box-Cox, Arcsine, Logit Transformations.

I. INTRODUCTION

THE Design and Analysis of Experiments (DOE) are efficient tools for the new product research and development, as well as process improvement, value-added product and cost reduction by applying statistical principles as experimental designs and analyses. Nowadays, DOE knowledge is necessary for researchers, engineers and related fields in such business and manufacturing sectors. One of manufacturer obstacles of these applications is the distinctness of understanding in such method since no manufacturing community consents to reveal any information, method or development process. Thereupon, the experimental design and analysis is as though impeding knowledge of that organization. However, any successfully occurred processes of the experimental designs and analyzes could be applied to solve other manufacturer's problems even similar or dissimilar. The tried knowledge, either normal or special process is able to utilize, facilitate and spread to public that cause the assistance in local manufacturers. Thus, this will be numerously advantageous if we could instruct the know-how from one to another, without any detriment or affection took place to the inventor.

In consideration of concealing the natural variables and responses to protect the mercantile secret, the data transformation is the key to accomplish, likewise, convince the inventors to reveal their valuable know-how to public. Meanwhile, the other organizations could apply this design and analysis experiment to appropriate their own process eventually. Data transformations are the applications of a mathematical modification to the values of variables and responses. There are a great variety of possible data transformations, from adding constants to multiplying, squaring or raising to a power, converting to logarithm scales, inverting and reflecting, taking the square root of the values, and even applying trigonometric transformations such as sine wave transformations for the sake of various coded in various task.

There are several researches related to the data transformation techniques since Freeman and Tukey [1] presented the data transformation for Poisson and binomial data by using the squared root and arcsine transformations. Afterwards, Box and Cox [2] presented the classical one of analysis of transformations via the lambda (λ) selection method for power transformation. John and Draper [3] presented an alternative family of transformation which is called the modulus transformation. It specially deals with the non-normal symmetric distribution with long tails. Kirisci, et al [4] used the simulation technique to study in multivariate statistical area that relies on the assumption of multivariate normality. Their research is about the effects of skewed and leptokurtic multivariate data on type I error and power of Hotelling's T-squared when applying the Box-Cox transformation to the data. The results indicate that even when variance-covariance matrices and sample sizes are equal, small to moderate changes in power still can be observed.

Besides the Box-Cox transformation, another option for normalize data which is positively skewed, often used when measuring reaction times, is the Ex-Gaussian distribution. It is a combination of the exponential and normal distribution. Olivier and Norberg [5] compared between Box-Cox transformation and Ex-Gaussian distribution when data is positively skewed. The numerical results demonstrate that

This work was supported by the Higher Education Research Promotion and National Research University Project of Thailand, Office of the Higher Education Commission. The authors wish to thank Thammasat University, Thailand for the financial support.

K. Chinda is a master student with the Industrial Statistics and Operational Research Unit (ISO-RU), Department of Industrial Engineering, Faculty of Engineering, Thammasat University, 12120, Thailand (phone: (662)564-3002-9; Fax: (662)564-3017; e-mail: prod_mbt@yahoo.co.th).

P. Luangpaiboon is an Associate Professor, ISO-RU, Department of Industrial Engineering, Faculty of Engineering, Thammasat University, 12120, Thailand (phone: (662)564-3002-9; Fax: (662)564-3017; e-mail: lpongch@engr.tu.ac.th).

Box-Cox transformation is simpler to apply and easier to interpret than the Ex-Gaussian distribution. Duran [6] has studied the use of arcsine transformation in the analysis of variance (ANOVA) when the data follow a binomial distribution. The Monte Carlo simulation technique was used to generate raw data. The results suggest that the transformed analyses do not always result in better type I error. In some cases they lose power and this provides some evidences to discourage the routine application of the arcsine transformation in ANOVA. Cordeiro and Andrade [7] have introduced a class of transformed symmetric models that extend the Box-Cox transformation model to more general symmetric models. This method is able to deal with all symmetric continuous distributions with a possible non-linear structure for the mean and enables the fitting of a wide range of models to several data types. They claimed that the proposed methods offer more flexible alternatives to Box-Cox or other existing procedures.

Recently, researches about coded data in experimental designs, researcher found that any transformation in such power by λ of Box-Cox, arcsine or logit transformations frequently point out the purpose of data improvement to be conformed with the assumption of "*Normally and Independently Distributed with equal variance or NID(0,σ^2)*" and apply to parametric tests afterwards. While the data transformation in our purpose, the natural variables data were in accordance with normality assumption, besides, the result of transformation have got to be tough for decoding, still be in accordance with normality assumption and the testing result or test statistic must be identical as before coding [8-14]. These issues are summarized in Table I.

TABLE I
A COMPARISON OF THE CODING APPEARANCE IN THE DIFFERENT PURPOSES

Data	Former purpose	Research Purpose
Natural data	Non-conform with *NID(0,σ^2)*	Conform with *NID(0,σ^2)*
Coded data	Conform with *NID(0,σ^2)* Easy to decode	Conform with *NID(0,σ^2)* Difficult to decode
F statistic	Up to transformation and no need to be the same as un-coded data	Need to be the same as un-coded data

This research takes three types of, Box-Cox, arcsine and logit transformations in the completely randomized design for a single factor, to study how the transformed variable such as the parameter λ of Box-Cox involves to F-statistic, the appropriate level for each specific research purpose, how to choose the proper variable levels. In addition, the influences of the number of replicates, the effect size and data dispersion against the efficiency of each transformation are also determined.

II. THEORETICAL ASPECTS

A. Experimental Design

Completely Randomized Design (CRD) or one way-ANOVA is the single factor design and analysis experiment, to compare more than two sets of the population mean, where a is the number of treatments [15]. The hypothesis of CRD is shown as

$$H_0: \mu_1 = \mu_2 = \cdots = \mu_a \qquad (1)$$
$$H_1: \mu_i \neq \mu_j, \text{ at least 1 pair}$$

Linear statistical model of the CRD is given below.

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij} \begin{cases} i = 1,2,\dots,a \\ j = 1,2,\dots,n \end{cases} \qquad (2)$$

where y_{ij} is the *ij* th observation, μ is the mean of population, τ_i is the *i* th treatment effect and ε_{ij} is the *ij* th random error or experimental residual. One way-ANOVA is the upper one tail test that will reject the null hypothesis when $F_0 > F_{\alpha,a-1,N-a}$, where *N* is the total number of experiments.The ANOVA categorized by its source of variation is shown in Table II, where *N* is the total number of experimental runs.

TABLE II
ANOVA TABLE FOR CRD

Source of Variation	Sum of Squares	df	Mean Squares	F_0
Between Treatments	$SS_{treatment}$	$a-1$	$MS_{treatment} = \frac{SS_{treatment}}{a-1}$	$F_0 = \frac{MS_{treatment}}{MS_E}$
Within Treatments	SS_E	$N-a$	$MS_E = \frac{SS_E}{N-a}$	
Total	SS_T	$N-1$		

B. Data Transformation

Box and Cox Transformation

In 1964, Box and Cox [2] presented the statistical and mathematical procedures to improve the data that in-accordance with *NID(0,σ^2)* by power transformation in order to improve the normality and constant variance of residuals. Box-Cox transformation benefits the optimal lambda calculation for the power transformation to reach the assumption. The transform equation is shown below.

$$y^{(\lambda)} = \begin{cases} \frac{y^{\lambda}-1}{\lambda \dot{y}^{\lambda-1}}\,;\ \lambda \neq 0 \\ \dot{y}\ln y\ ;\ \lambda = 0 \end{cases} \qquad (3)$$

where $\dot{y} = \sqrt[n]{\prod y}$ is the geometric mean of the observations, *n* is total number of replicates. Box-Cox method selects the optimal lambda which causes the lowest sum square error or $SS_E(\lambda)$ or the lowest pooled standard deviation or S_p (Fig. 1). The value of S_p is illustrated in an equation below.

$$S_p = \sqrt{\frac{\Sigma_i \Sigma_j \left(y_{ij}^{(\lambda)} - \overline{y_i^{(\lambda)}}\right)^2}{\Sigma_i (n_i - 1)}} \tag{4}$$

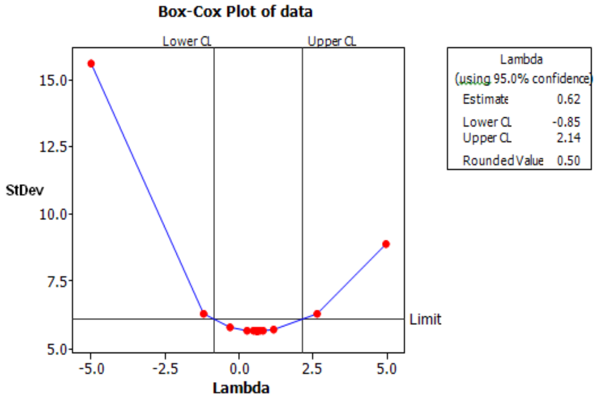

Fig. 1 The relationship of the pooled standard deviation and its lambda via Minitab program

The two most common methods for transforming percent, proportion and probabilities are the arcsine and logit transformations. [8] In both cases, the percentage should firstly be changed to the proportion by dividing the percentage by 100. These transformations are applicable only to the percentages that lie between 0 and 100. They should not be used in the case of the percent increase which can give values greater than 100%.They frequently use when there are a number of proportions close to 0 and/or close to 1.The transformations will stretch out the proportions that are close to 0 and 1 and compress the proportions close to 0.5.

Arcsine Transform

Sometimes called an angular transformation, the arcsine transform equals the inverse sine of the square root of the proportion or

$$y = arcsine(\sqrt{p}) \tag{5}$$

where p is the proportion and y is the transformation result. The result may be expressed either in degrees or radians.

Logit Transform

A logit is defined as the logarithm of the odds. If p is the probability of an event, then $(1 - p)$ is the probability of not observing that event and the odds of the event are $p/(1 - p)$. The logit transformation is most frequently used in logistic regression and for fitting linear models to categorical data (log-linear models). The logit transformation is undefined when $p = 0$ or $p = 1.0$. This is not a problem with either of the two above-named techniques because the logit transformation is applied to a predicted probability which can be shown to always be greater than 0 and less than 1.0. Hence, the logit transformation is

$$y = logit(p) = log\left(\frac{p}{1-p}\right) \tag{6}$$

As arcsine and logit transformations are required for proportion. In this research, the simulated data were the real numbers, so it will be defined as y/Ω, where y is simulated data and Ω is the transformed variable.

III. Methodology

A. Data Simulation

The data is simulated from the linear statistical model. There are five treatments, three factors with three levels with the replicates (rep) of 2, 5 and 10, the levels of the treatment effects (τ) at 1.02, 1.05 and 1.10 times from the precedent treatment and the standard deviation (σ) levels at 1.0, 3.0 and 5.0. There are totally 27 scenarios to be tested.

B. Experiments

The data as defined conditions were simulated by Minitab program while the Matlab program was encoded [16]. Then, both programs were used to run the CRD experiments and F_0 of natural data (F_n) were then collected. Afterwards, the achieved data were coded with Box-Cox, arcsine and logit transformations. A feasible range of λ is from -5 thru 5 for Box-Cox and a feasible range of Ω for arcsine and logit is from 0 thru 500. In the meanwhile, F_0 of coded data (F_c) was calculated and then compared the result between Matlab and Minitab to ensure the correction of encoded program. Repeat the whole sequential procedures for 100 times and finally analyze the result of experiments.

IV. Result and Discussion

The experimental results of 27 scenarios which were simulated from Minitab program were gathered. The result of F_n optimal λ and F_c of Box-Cox transformation as shown in Table III. Notice that the F_c is different from F_n in all scenarios, mostly nearby F_n but there are large differences in the scenarios of 4, 5, 7, 8 and 9. The investigation showed that the tendency of significant differences between F_c and F_n at 2 replicates, likewise, a bit difference at 5 and 10 replicates as illustrated in Fig. 2, 3 and 4, respectively.

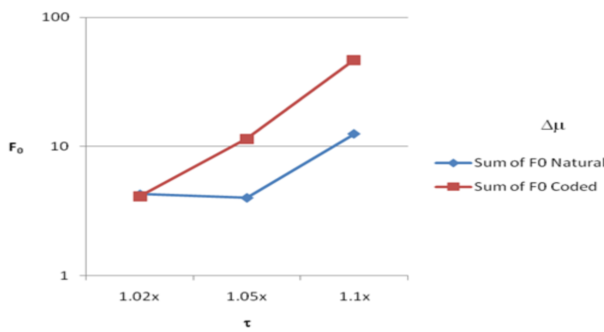

Fig. 2 Comparative study of F_c and F_n with 2 replicates

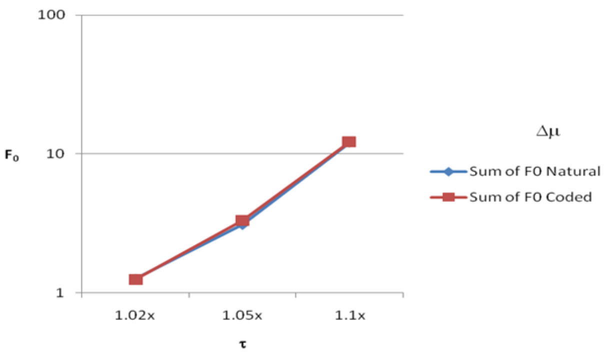

Fig. 3 Comparative study of F_c and F_n with 5 replicates

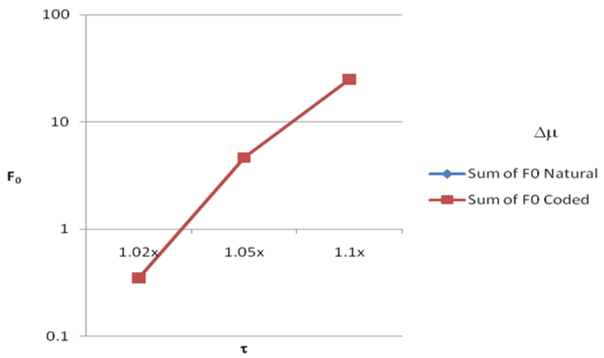

Fig. 4 Comparative study of F_c and F_n with 10 replicates

TABLE III
RESULT OF BOX-COX TRANSFORMATION ON THE CRD

No.	Rep	τ	σ	Natural data F_n	P-value	λ	Coded data F_c	Pvalue
1	2	1.02x	1.0	10.61	0.012	4.12	10.77	0.011
2	2	1.02x	3.0	4.27	0.072	0.43	4.09	0.077
3	2	1.02x	5.0	0.38	0.818	-1.56	0.30	0.870
4	2	1.05x	1.0	33.03	0.001	-5	82.62	0.000
5	2	1.05x	3.0	4.01	0.080	-4.33	11.46	0.010
6	2	1.05x	5.0	0.69	0.628	3.93	0.58	0.690
7	2	1.1x	1.0	355.72	0.000	3.43	526.27	0.000
8	2	1.1x	3.0	12.58	0.008	-5	47.02	0.000
9	2	1.1x	5.0	1.11	0.443	3.68	2.00	0.233
10	5	1.02x	1.0	3.70	0.021	3.8	3.77	0.019
11	5	1.02x	3.0	1.26	0.318	2.41	1.24	0.327
12	5	1.02x	5.0	0.30	0.878	1.36	0.30	0.872
13	5	1.05x	1.0	16.07	0.000	1.3	16.10	0.000
14	5	1.05x	3.0	3.10	0.039	-0.14	3.30	0.031
15	5	1.05x	5.0	1.76	0.177	0.62	1.68	0.195
16	5	1.1x	1.0	78.02	0.000	0.79	77.75	0.000
17	5	1.1x	3.0	12.12	0.000	1.65	12.20	0.000
18	5	1.1x	5.0	12.19	0.000	1.79	13.60	0.000
19	10	1.02x	1.0	3.03	0.027	-1.8	2.95	0.030
20	10	1.02x	3.0	0.35	0.842	0.94	0.35	0.841
21	10	1.02x	5.0	2.25	0.079	0.77	2.21	0.083
22	10	1.05x	1.0	78.86	0.000	1.62	79.05	0.000
23	10	1.05x	3.0	4.67	0.003	1.16	4.66	0.003
24	10	1.05x	5.0	3.28	0.019	0.81	3.29	0.019
25	10	1.1x	1.0	169.82	0.000	-0.95	193.22	0.000
26	10	1.1x	3.0	25.00	0.000	1.07	25.09	0.000
27	10	1.1x	5.0	11.10	0.000	0.93	11.07	0.000

After a consideration on Matlab of the numerical results for scenarios 4, 5, 7, 8 and 9, the relationships between F_0 and λ for the scenario 4 is illustrated in Fig. 5.

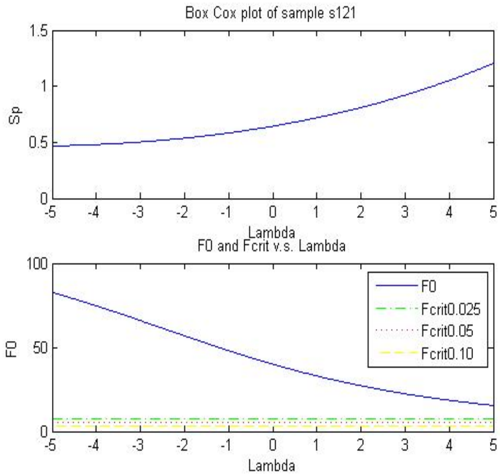

Fig. 5 The relationship of F_0 and λ for the scenario 4

The optimal λ from the Box-Cox method (at lowest s_p) is at -5 which results F_c = 82.62, that the large difference from F_n (33.03) is at λ = 1. In this situation, noticed that F statistic increases while λ decreases. Anyway, F statistic lies above F_{crit} at every significant level through the range of $\lambda$from -5 thru 5. Thus, if we need only the result of hypothesis test (F_c is no need to equal F_n), it is free to select any λ from -5 thru 5 without any change in result of hypothesis testing at all.

As per scenario 5, illustrated in Fig.6, the optimal λ from Box-Cox method is -4.33 which results F_c = 11.46, that the large difference from F_n (4.01) at λ = 1. As well as the scenario 4, F statistic increases while λ decreases but this scenario, λ selection depends on the significance level of testing due to the result of natural data testing (λ =1) reject the null hypothesis at significant level (α) = 0.10, however, neither at α = 0.05 nor α = 0.025. Therefore, at α = 0.10, λ levels have got to be not over than 1.75 indeed, while λ must be less than -0.42 and 2.21 at α = 0.05 and α = 0.025 respectively. Under this mentioned condition, the result of hypothesis testing would certainly remain as ever.

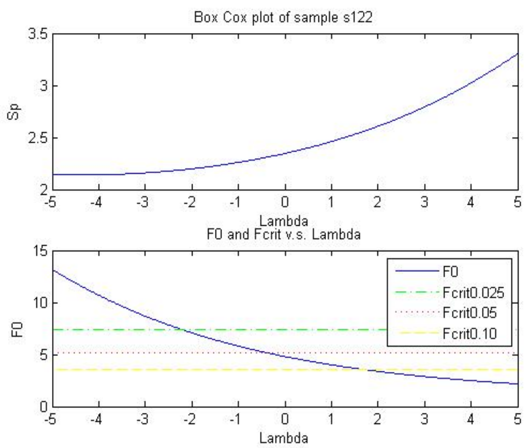

Fig. 6 The relationship of F_0 and λ for the scenario 5

In the scenario 7 as illustrated in Fig. 7, the optimal λ from Box-Cox method is 3.43 which results F_c = 526.27 that big different from F_n (355.72) at λ =1. F statistic increase following to λ and lies above F_{crit} at every significant level through the range of λ from -5 thru 5. As scenario 4, if we need only the result of hypothesis test (F_c is no need to equal F_n), it is free to select any λ from -5 thru 5 without any change in result of hypothesis testing at all.

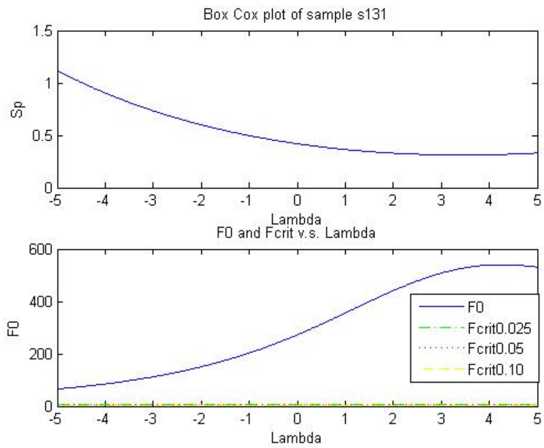

Fig. 7 The relationship of F_0 and λ for the scenario 7

The scenario 8, illustrated in Fig. 8, the optimal λ from the Box-Cox method is at -5 which results F_c = 47.02 that the large difference from F_n (12.58) at λ =1. F statistic increases while λ decreases, in this case, F_n rejects the null hypothesis at every significance level but F statistic lines cross the F_{crit} α = 0.025 line at λ = 3.51. By this reason, λ selection at α = 0.025, λ must be not over than 3.51, likewise, at α = 0.05 and α = 0.10, all range of λ could be selected by free without any change on result of hypothesis testing.

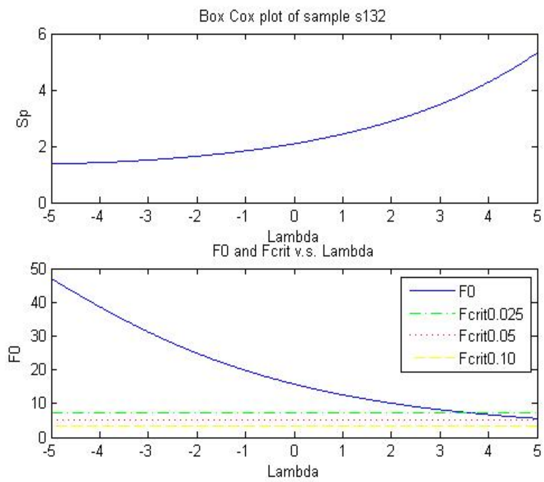

Fig. 8 The relationship of F_0 and λ for the scenario 8

The scenario 9, illustrated in Fig. 9, the optimal λ from the Box-Cox method is at 3.68 which results F_c = 2.00 that the large difference from F_n (1.11) at λ = 1. In this case, F statistic lies below F_{crit} at every significance level. Neither, F_c nor F_n reject null hypothesis that it is free to select any λ from -5 thru 5 without any change in results of hypothesis testing at all.

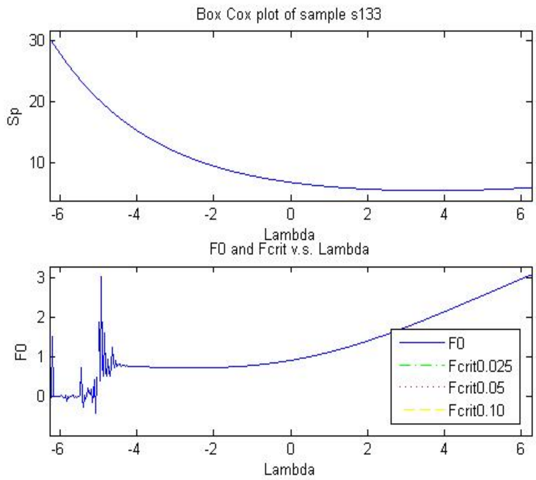

Fig. 9 The relationship of F_0 and λ for the scenario 9

Besides the above scenarios, there is remarkable arisen in such scenario 24 as shown in Fig. 10, the optimal λ from the Box-Cox method is at 0.81 which results F_c = 3.29 and the natural data result F_n = 3.28 at λ = 1. The Box-Cox transformation is much appropriate for purpose since F_c is nearby F_n that because the optimal λ gets close 1, where coded data also get close natural data. In this scenario, the more F_c of Box-Cox gets close to F_n, the more coded data gets close to natural data. In other words, Box-Cox transformation never cause F_c equal to F_n, except λ = 1 or no data transformation.

Thus, the other choice in the λ selection is considered from the significance level. In this case, at α = 0.025, λ should be within the range of -1.02 thru 2.19. As α = 0.05, λshould be within the range of -2.55 thru 3.91, and α = 0.10, λ should be more than -3.89, in order to be still during the hypothesis testing.

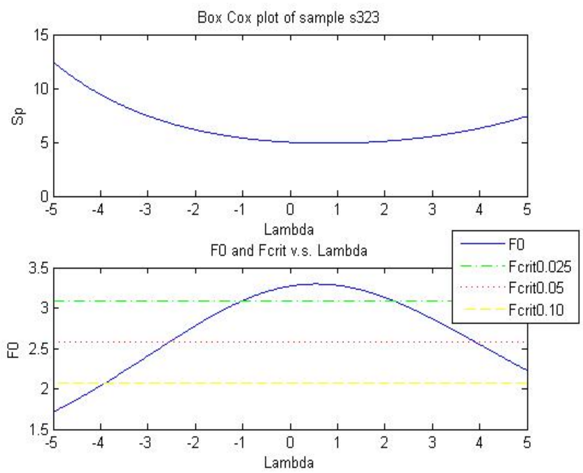

Fig. 10 The relationship of F_0 and λ for the scenario 24

International Journal of Mechanical, Industrial and Aerospace Sciences
ISSN: 2517-9950
Vol:6, No:12, 2012

According to the data shape of five treatments of natural data (a), Box-Cox transformation data (b) which takes the minus λ *(or* $-\lambda$*)* reveals the characteristic of the reciprocal transformation and causes the coded data shape invert with the natural data (Fig. 11).

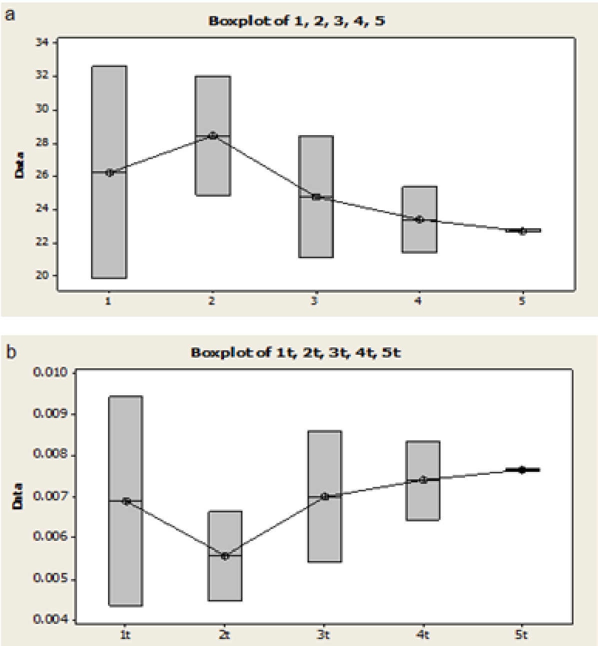

Fig. 11 The comparative results of Box-Cox transformation with -λ

A comparison on the ANOVA results of arcsine and logit transformations with natural data, the numerical result of F_n optimal Ω and F_c are shown in Table IV.

TABLE IV
ARCSINE AND LOGIT TRANSFORMATION RESULTS ON THE CRD

Data No.	Natural Variables	Coded: Arcsine		Coded: Logit	
	F_n	Ω	F_c	Ω	F_c
1	10.61	217.7	10.61	28.4369	10.61
2	4.27	49.8478	4.27	49.7710	4.27
3	0.38	50.333	0.38	50.3435	0.38
4	33.03	59.176	33.03	59.1711	33.03
5	4.01	50.8373	4.01	50.8524	4.01
6	0.69	45.0502	0.69	44.9611	0.69
7	355.72	55.5191	355.72	55.4767	355.72
8	12.58	67.3241	12.58	67.3469	12.58
9	1.11	55.3916	1.11	55.4074	1.11
10	3.70	47.9325	3.70	47.8978	3.70
11	1.26	47.722	1.26	47.8263	1.26
12	0.30	42.6066	0.30	43.3029	0.30
13	16.07	49.06	16.04	51.20	16.01
14	3.10	58.8224	3.10	58.8269	3.10
15	1.76	59.4152	1.76	59.3233	1.76
16	78.02	55.0775	78.02	54.62	77.95
17	12.12	57.22	12.03	58.48	11.93
18	12.19	52.4978	12.19	52.1885	12.19
19	3.03	54.7882	3.03	54.7876	3.03
20	0.35	58.3848	0.35	58.478	0.35
21	2.25	52.9409	2.25	52.9231	2.25
22	78.86	46.916	78.86	45.10	78.84
23	4.67	55.6031	4.67	55.6307	4.67
24	3.28	72.4766	3.28	162.8615	3.28
25	169.82	68.5424	169.82	68.5809	169.82
26	25.00	47.9499	25.00	50.62	24.86
27	11.10	59.50	10.98	60.96	10.85

Note that the results of F_c from arcsine and logit transformations mostly be exactly equal to F_n, except scenarios 13, 17 and 27 for the arcsine transformation, and scenarios 13, 16, 17, 22, 26 and 27 for the logit transformation that be a bit difference but still keep in the same result as hypothesis testing anyway.The reason why arcsine and logit transformations are able to transform data while keeping the same or similar F statistic, after consideration, it was found that the relationship of F statistic and Ω in arcsine and logit transformations could be categorized in three types of F_c line intersect F_n (A), F_c line lies above F_n(B) and F_c line lies below F_n (C). As per (A) type, it is probable to have one or two intersection points which are the optimal values of Ω as shown in Fig. 12 and 13 are the samples of this type from scenarios 4 and 7, respectively.

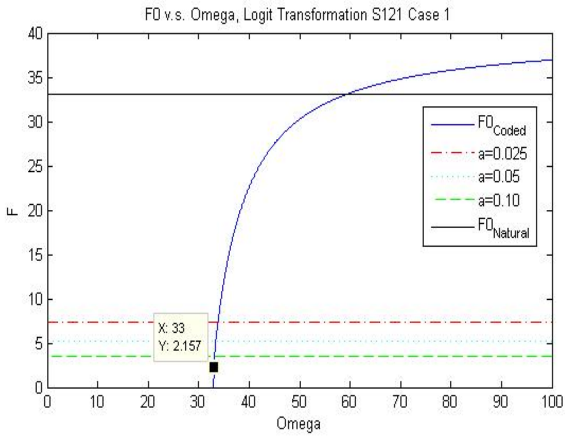

Fig. 12 Logit transformation of the (A) type with only one intersection point

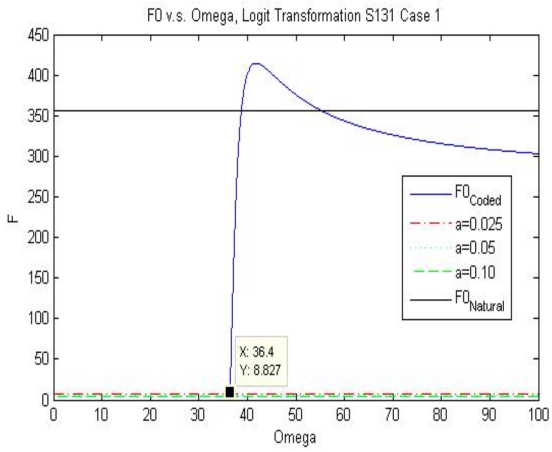

Fig. 13 Logit transformation of the (A) typewith 2 intersection points

The (B) typewhich F_c line lies above F_n, hence the optimal Ω for this type is at the lowest point of the graph that F_c gets closest to F_n. Fig. 14 shows the sample of this type by the scenario 1 that the lowest point is at Ω = 217.7.

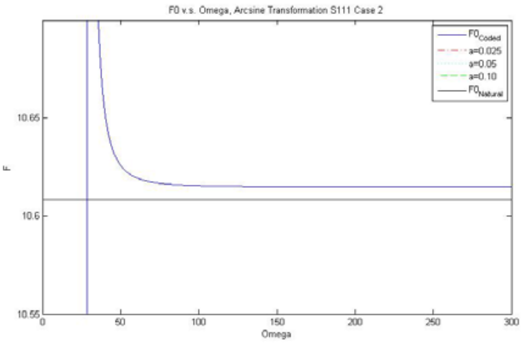

Fig. 14 Arcsine transformation of the (B) type

The (C) type which F_c line lies below F_n, hence the optimal Ω for this type is at the highest point of the graph that F_c gets closest to F_n. Fig. 15 shows the sample of this type by the scenario 27 that the highest point is at $\Omega = 59.50$.

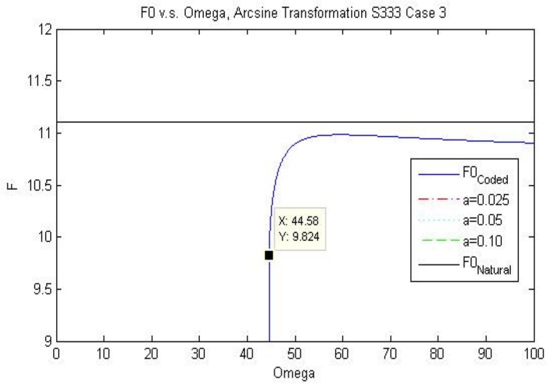

Fig. 15 Arcsine transformation of the (C) type

Therefore, refer to Table IV, $F_c = F_n$ situation is the result from the (A) type while the other situations that F_c is a bit different from F_n are the results from the (B) and (C) types. As per (A) type, when there are 2 intersection points, researchers select the second intersection point as the optimal Ω due to the matter of slope, since the first intersection point is usually high slope (such as the logit transformation of the (A) type in scenario 1 with the first intersection slope ~ 143.71). That causes the extremely change in F_c, despite a bit change in Ω. On the other hand, the slope of the second intersection point is normally less than the first. That is to say, if we select the Ω from the first intersection point, we must define the number of decimal precisely, at least 4 digits after the decimal point. However, the second intersection point required only 1 or 2 digit of decimal sufficiently.

Afterwards of the 100 sequential experiments in every scenario, the numerical results indicated that the arcsine and logit transformations are more effective than Box-Cox (BC) transformation, regarding of the accuracy, considerate from mean of F statistic difference ($\overline{\Delta F} = average(F_c - F_n)$)and the precision, considerate from standard deviation of F statistic different ($SD_{\Delta F} = SD(F_c - F_n)$). In addition, the number of null hypothesis rejected of the arcsine (AS) and logit (LG) transformations are equal to natural data (*N*) in almost scenarios, except only scenario 9 which the logit transformation is less than natural data just one time at the significance level of 0.05 and 0.10. Fig 16 shows the sample of numerical results from the scenario 1. Table V shows summary results *of* $\overline{\Delta F}$ and $SD_{\Delta F}$ for the whole 27 scenarios and Table VI shows the number of null hypothesis rejected from the scenario 9.

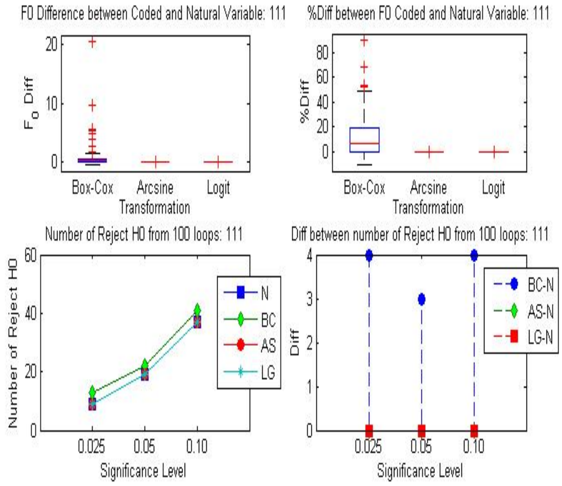

Fig. 16 The numerical results from the sequential experiments from the scenario 1

TABLE V
THE SUMMARY RESULTS OF $\overline{\Delta F}$ *AND* $SD_{\Delta F}$ FOR THE WHOLE 27 SCENARIOS

Data No.	Box-Cox		Arcsine		Logit	
	$\overline{\Delta F}$	$SD_{\Delta F}$	$\overline{\Delta F}$	$SD_{\Delta F}$	$\overline{\Delta F}$	$SD_{\Delta F}$
1	0.8	2.44	0	0	0	0
2	1.47	10.94	0	0.01	0	0.01
3	0.23	0.58	0	0.01	0	0.02
4	5.5	8.71	0	0.01	-0.01	0.02
5	0.92	3.01	0	0.01	0	0.02
6	1.61	8.95	0	0.02	0	0.03
7	34.28	90.94	-0.16	1.26	-0.36	2.53
8	6.31	18.95	-0.01	0.06	-0.03	0.1
9	1.11	2.93	-0.01	0.03	-0.02	0.08
10	0.18	0.38	0	0	0	0
11	0.07	0.19	0	0	0	0.01
12	0.05	0.24	0	0	0	0.01
13	1.71	2.82	-0.01	0.02	-0.02	0.05
14	0.27	1.11	0	0	0	0.01
15	0.09	0.28	0	0.01	0	0.02
16	11.66	18.2	-0.26	0.46	-0.61	1.01
17	0.63	1.53	-0.03	0.1	-0.08	0.21
18	0.21	0.47	-0.02	0.1	-0.06	0.21
19	0.12	0.31	0	0	0	0
20	0.05	0.15	0	0	0	0.01
21	0.02	0.12	0	0.01	0	0.01
22	1.54	2.73	-0.02	0.04	-0.04	0.09
23	0.08	0.24	-0.01	0.02	-0.01	0.05
24	0.04	0.14	0	0.01	0	0.02
25	7.15	10.8	-0.48	0.77	-1.24	1.66
26	0.55	1.18	-0.07	0.12	-0.17	0.25
27	0.36	1.09	-0.03	0.07	-0.09	0.16

TABLE VI
THE NUMBER OF NULL HYPOTHESIS REJECTED FROM THE SCENARIO 9

Scenario 9	H_0 reject				Diff from N		
Sig. level	N	BC	AS	LG	BC	AS	LG
α = 0.025	11	19	11	11	8	0	0
α = 0.05	24	32	24	23	8	0	-1
α = 0.10	41	46	41	40	5	0	-1

A consideration of the influence of replicate (*rep*), treatment effect (τ), and standard deviation (σ) against the efficiency of each transformation method, we made the hypothesis testing on result of 100 sequential experiments, for the whole 27 scenarios by defining 27 combinations of replicate (*rep*), treatment effect (τ), and standard deviation (σ) levels as the treatments of these experiments, the responses are F_c-F_n of each transformation. Because the residual of experiment is in-accordance with $NID(0,\sigma^2)$, we choose the non-parametric "Kruskal-Wallis test" in this purpose. Table VII shows the numerical results of Kruskal-Wallis test.

TABLE VII
THE NUMERICAL RESULTS FROM KRUSKAL-WALLIS TEST

Transform	Box-Cox		Arcsine		Logit	
Replicate	Median	Z	Median	Z	Median	Z
2	0.07125	13.59	0.000	4.19	0.000	5.05
5	0.01555	-3.29	0.000	0.01	0.000	0.00
10	0.00725	-10.30	0.000	-4.19	0.000	-5.05
P-value	0.000		0.000		0.000	
τ	Median	Z	Median	Z	Median	Z
1.02	0.01315	-5.8	0.000	6.12	0.000	8.80
1.05	0.01625	-0.6	0.000	2.16	0.000	2.55
1.1	0.02605	6.4	0.000	-8.28	0.000	-11.35
P-value	0.000		0.000		0.000	
σ	Median	Z	Median	Z	Median	Z
1.0	0.0265	6.45	0.000	-1.62	0.000	-1.89
3.0	0.0166	-1.34	0.000	0.04	0.000	0.14
5.0	0.01275	-5.12	0.000	1.58	0.000	1.76
P-value	0.000		0.183		0.108	

As per Box-Cox transformation, all of *rep*, τ, and σ reject the null hypothesis. The results indicate the data with larger replicates, smaller τ and larger σ should be more proper for this type of transformation. These conditions utilize the F_c results closer to F_n. This result is accordance with Figs. 2, 3 and 4, anyway, the non-parametric test is incapable to answer the question whether the interaction between each variable significantly affect against the response or not. About arcsine and logit transformations, almost factors reject the null hypothesis, except σ are unable to reject the null hypothesis. This is to say; only replicate and τ affect the efficiency of arcsine and logit transformations. However, noticed the median of replicates and τ show a very little effect, when compared with the Box-Cox transformation.

V. CONCLUSION

As mentioned above, data transformation for concealing the mercantile secret purpose. Researchers recommend to apply arcsine and logit transformations instead of Box-Cox transformation since these two methods could provide the result of F_c equal to F_n or at least similar. While Box-Cox transformation never cause F_c equal to F_n, except λ =1 or no data transformation. If we try to use λ close to 1 to obtain F_c get close to F_n, it causes the coded data get too close to natural data as well. However, if we need only the results of hypothesis test equal to natural variable (F_c has no need to be equal to F_n) Box-Cox transformation is acceptable when the optimal λ obtain from the range which results the same hypothesis test as natural variable, instead of the former methods.

ACKNOWLEDGMENT

This work was supported by the Higher Education Research Promotion and National Research University Project of Thailand, Office of the Higher Education Commission. The authors wish to thank the Faculty of Engineering, Thammasat University, Thailand for the financial support.

REFERENCES

[1] M.F. Freeman and J.W. Tukey, "Transformation Related to the Angular and the Square Root", The Annals of Mathematical Statistics, Vol. 10, 1939, pp. 247-253.
[2] G.E.P. Box and D.R. Cox, "An Analysis of Transformation", Journal of the Royal Statistical Society. Series B (Methodological), Vol. 26, No. 2, 1964, pp. 211-252.
[3] J.A. John and N.R. Draper, "An Alternative Family of Transformations", Journal of the Royal Statistical Society. Series C (Applied Statisyics), Vol. 29, No. 2, 1980, pp. 190-197.
[4] L. Kirisci, A.A. Al-Subaihi and R. Tarter, "Effects of the generalized Box-Cox transformation on Type I error rate and power of Hotelling's T^2", Journal of Statistical Computation and Simulation, Vol. 75, No. 3, March 2005, pp. 199-206.
[5] J. Olivier and M.M. Norberg, "Positively Skewed Data: Revisiting the Box-Cox Power Transformation", International Journal of Psychological Research, 2010, Vol. 3, No. 1, pp. 69-78.
[6] M.J. Duran, " The Use of the Arcsine Transformation in the Analysis of Variance when Data Follow a Binomial Distribution", Master Thesis, State Univ. of New York, College of Environmental Science and Forestry Syracuse, New York , May 1997.
[7] G.M. Cordeiro and M.G. Andrade, "Transformed symmetric models", International Journal of Statistical Modelling, Vol. 11(4), 2011, pp. 371-388.
[8] "General Statistics *Transformations*" Code No. PSYC-5741 (4), Department of Psychology and Neuroscience, University of Colorado Boulder, USA (Online) Available : http://psych.colorado.edu/
[9] M.S. Bartlett, "The use of Transformation", Biometric Bulletin Vol. 3, 1947, pp. 39-52.
[10] J.L. Rasmussen and W.P. Dunlap, "Dealing with Nonnormal Data: Parametric Analysis of Transformed data VS Nonparametric Analysis", Educational and Psychological Measurement Journal Winter 1991, Vol. 51, pp. 809-820.
[11] Y. Guan, "Variance stabilizing transformations of Poisson, binomial and negative binomial distributions", Statistics and Probability Letters, Vol. 79, 2009, pp. 1621-1629.
[12] F.J. Anscombe, "The Transformation of Poisson, binomial, negative binomial data", Biometrika Journal, Vol. 35, 1948, pp. 246-254.
[13] M.S. Bartlett, "The square root transformation in the analysis of variance", Journal of the Royal Statistical Society, Vol.3, 1936, pp. 68-78.
[14] G.E.P. Box and D.R. Cox, "An analysis of transformation Revisited", Journal of American Statistical Association, Vol. 77, 1982, pp. 177-182.

[15] P.A. Bromiley and N. A. Thacker, "The effects of an arcsine square root transform on a binomial distributed quantity", TINA memo, 2002, pp. 2002-007.
[16] D.C. Montgomery, *Design and Analysis of Experiments*, Hoboken, NJ: John Wiley & Sons, Inc., 2001.
[17] S.J. Chapman, *MATLAB Programming for Engineers*, Pacific Grove, CA: Wadsworth, 2002.

K. Chinda is a master student in the Industrial Statistics and Operational Research Unit (ISO-RU), the department of Industrial Engineering at Thammasat University, Thailand. He graduated his Bachelor from Burapha University, Thailand. His research interests consist of industrial statistics, quality improvement and response surface methodology.

P. Luangpaiboon has been a lecturer, and Associate Professor, in the Industrial Statistics and Operational Research Unit (ISO-RU), the department of Industrial Engineering at Thammasat University, Thailand since 1995. He graduated his Bachelor (1989-1993) and Master Degrees (1993-1995) in Industrial Engineering from Kasetsart University, Thailand and Ph. D. (1997-2000) in Engineering Mathematics from Newcastle upon Tyne, England. He is a member of International Association of Computer Science and Information Technology (IACSIT) and International Association of Engineers (IAENG). His research interests consist of meta-heuristics, optimisation, industrial statistics, the design and analysis of experiments and response surface methodology. He received Kasetsart University Master Thesis Award in 1995 (Dynamic Process Layout Planning), Certificate of Merit for The 2009 IAENG International Conference on Operations Research (A Hybrid of Modified Simplex and Steepest Ascent Methods with Signal to Noise Ratio for Optimal Parameter Settings of ACO), Best Paper Award for the Operations Research Network Conference 2010 (An Exploration of Bees Parameter Settings via Modified Simplex and Conventional Design of Experiments), Certificate of Merit for The 2011 IAENG International Conference on Operations Research (Bees and Firefly Algorithms for Noisy Non-Linear Optimisation Problems) and Best Student Paper Award for The 2011 IAENG International Conference on Industrial Engineering (Simulated Manufacturing Process Improvement via Particle Swarm Optimisation and Firefly Algorithms). He was a local chair and editor of the 4th International Conference on Applied Operational Research (ICAOR'12) and Lecture Notes in Management Science (LNMS) Volume 4 July 2012, respectively.