Statistika, Vol. 7 No. 1, 45 – 50
Mei 2007

A Modification of Box-Cox Transformation

WAN MUHAMAD AMIR BIN WAN AHMAD[A)], NYI NYI NAING[B)], MOHD TENGKU ARIFF RAJA HUSSEIN[C)]

[A)] Jabatan Matematik, Fakulti Sains dan Teknologi, Malaysia, Universiti Malaysia Terengganu UMT, 21030 Kuala Terengganu, Terengganu Malaysia.
[B)] Unit Biostatistics and Research Methodology, School of Medical Sciences, Universiti Sains Malaysia, USM, Healthy Campus, 16150 Kubang Kerian, Kelantan, Malaysia.
[C)] Department of Community Medicine, School of Medical Sciences, Universiti Sains Malaysia, USM, Healthy Campus, 16150 Kubang Kerian, Kelantan, Malaysia.

ABSTRACT

Data screening is the most important technique to check the nature of the data. One of the methods to screen the data is the transformation method. There are a lot of transformations but the famous one is the Box-Cox Transformation. The purpose of this study is to find the alternative way of transformation by making some modification of the original Box-Cox formula. As we know, the Box-Cox transformation method only transform the data if and only if the data in a positive values. If the values in the analysis are negative, the formula of the Box-Cox cannot be used. This problem always occur when the researcher want to transform the data which the values is negative. In this methodology section, we will build an alternative method to solve the problem if the data in the study are negative.

Keywords : Box-Cox transformation, modified of Box-Cox transformation

1. Introduction

The scale on which a variable is measured may not be the most appropriate for statistical analysis or describing variation and may even hide basic characteristics of the data. Moreover, statistical techniques are based on assumption and the validity of the results obtained from them depends on the assumed conditions being satisfied. In practice, however, these assumptions are often not even approximately satisfied. For example, when the variable of interest has support on the positive real line and is positive skewed, normal theory results are not directly applicable. To model such non-normal data, one can transform the observation to meet the necessary distributional assumptions.

In general, a re-expression of the data may uncover some of its basics properties and allow for the valid used of many powerful standard statistical method s on the transformed observations. For example, when the assumption of the analysis of variance are not satisfied by the measurement on the original scale, a power transformation of the measurement makes the assumptions more tenable. Tukey (1957) regards transformation as re-expressions of the data and basic tools of data analysis and statistical inference.

Moore and Tukey (1954) initiated the study of family of transformations. They introduced the simple family of transformations

$$\rho_{\lambda_1,\lambda_2}(X) = \begin{cases} (X+\lambda_1)^{\lambda_2}, & 0 \neq \lambda_2 \leq 1, \\ \log(X+\lambda_1), & \lambda_2 = 0, \end{cases} \quad \text{for } X+\lambda_1 > 0\,.$$

In a practical application, the selection of λ_1 and λ_2 depends on what is desired by transformation. Moore and Tukey (1954) select λ_1 and λ_2 so that interactions are not significant and additivity holds in two-way layouts. Later Box and Cox (1964) modified the simple family of transformation and considered the following transformation:

$$\rho_\lambda(X)=\begin{cases}\dfrac{X^\lambda-1}{\lambda}, & \lambda\neq 0,\\ \log(X), & \lambda=0,\end{cases}\quad \text{for } X>0.$$

To accommodate negative observations, Box and Cox (1964) further proposed the following shift-transformation:

$$\rho_{\lambda_1,\lambda_2}(X)=\begin{cases}\dfrac{(X+\lambda_1)^{\lambda_2-1}}{\lambda_2}, & \lambda_2\neq 0,\\ \log(X+\lambda_1), & \lambda_2=0,\end{cases}\quad \text{for } X+\lambda_1>0.$$

It is sufficient to substitute a convenient value for λ_1 and use the Box-Cox transformation on the shifted value of $X+\lambda_1$. If X has a minimum, then $X+\lambda_1$ can be made positive by the proper choice of λ_1. See Atkinson (1985) for a detailed discussion.

To obtain approximate normality from symmetric long-tailed distributions, John and Draper (1980) propose another family called the family of modulus transformations

$$\rho_\lambda(X)=\begin{cases}\operatorname{sgn}(X)\{[(|X|+1)^\lambda-1]/\lambda\}, & \lambda\neq 0,\\ \operatorname{sgn}(X)\log(|X|+1), & \lambda=0,\end{cases}$$

Where $\operatorname{sgn}(\cdot)$ is the sign function. The modulus family is monotonic, continuous in λ and is applicable in the presence of negative data. Note that when the observations are all positive, the transformation reduces to the Box-Cox power transformation family. The basic idea is to apply the same power transformation to both tails of a distribution symmetric about zero. Another modification to the Box-Cox transformation by Bickel and Doksum (1981) is

$$P_{BD\lambda}(X)=[X^\lambda\operatorname{sgn}(X)-1]/\lambda,\ \lambda>0,\ X\in(-\infty,\infty)$$

for an unlimited support.

Three well known classes of transformation are rank transformations (Lehmann, 1975). Transformation to symmetry, and transformation to normality (Box-Cox 1964). Here a rank Transformation induces uniformity, a symmetrizing transformation brings about symmetry and a normalizing transformation induces normality. Transformations to normality have played a prominent role in the analysis of non-normal data. Statisticians have an arsenal of statistical method at their disposal when the underlying of the observations is normal. Such powerful normal theory based procedures can be applied to the transformed data. Miller (1986) discussed transformation along with nonparametric technique and robust estimation as method of correcting for non-normality. Chen and Loh (1992) study the gain in the efficiency of tests following a Box-Cox transformation to a more suitable case scale. Hinkely and Runger (1984) provide an overview of the analysis of transformed data. An extensive bibliography on transformations is compiled by Hoyle (1973). In this work, we assume $X>0$ and the Box–Cox transformation is define as

$$Y=\rho_\lambda(X)=\begin{cases}\dfrac{X^\lambda-1}{\lambda}, & \lambda\neq 0\\ \ln X, & \lambda=0\end{cases}\tag{1}$$

Here X is a positive random variable and λ is the transformation parameter.

2. A Modified of Box-Cox Transformation

In this section we will introduced the alternative method of transformation. This alternative method is based on Box-Cox method but it's involved some of modification. Box- Cox transformation, for example is concave in x for $\lambda<1$ and convex in x for $\lambda>1$. How ever

$$\rho_\lambda(X)=\{\operatorname{sgn}(X)|X|^\lambda-1\}/\lambda,\quad \lambda>0$$

changes from convex to concave as x changes sign so it is not to be recommended when data that can be positive or negative skewed. John and Draper (1980) and Burbidge, Magee & Robb (1988) studied specific cases of other convex-to-concave transformations.

Now we first consider a modified modulus transformation which has different transformation parameters on the positive line

$$\xi(\lambda_+, x) = \rho_\lambda(X) = \begin{cases} \dfrac{(X+1)^{\lambda+} - 1}{\lambda_+} & (x \geq 0, \lambda_+ \neq 0), \quad (2) \\ \log(X+1) & (x \geq 0, \lambda_+ = 0), \quad (3) \end{cases}$$

and for the negative line

$$\xi(\lambda_-, x) = \rho_\lambda(X) = \begin{cases} -\left\{\dfrac{(-X+1)^{\lambda-} - 1}{\lambda_-}\right\} & (x < 0, \lambda_- \neq 0), \quad (4) \\ \log(-X+1) & (x < 0, \lambda_- = 0), \quad (5) \end{cases}$$

Next we find first, second till to k derivation for (1) to (5) with the condition of the second derivative $\partial^2 \psi(\lambda_+, \lambda_-, x) / \partial x^2$ be continuous at $x = 0$.

The first and the second derivative of the (2) is given by

$$\xi(\lambda_+, x) = \frac{(x+1)^{\lambda+} - 1}{\lambda_+}$$

First derivative

$$\xi'(\lambda_+, x) = \frac{\lambda_+ (x+1)^{\lambda_+ - 1}}{\lambda_+}$$

$$= \lambda_+ x^{\lambda_+ - 1}$$

Second derivative

$$\xi''(\lambda_+, x) = \lambda_+ - 1(x+1)^{\lambda_+ - 2} |_{x=0}$$

$$= \lambda_+ - 1 \quad (6)$$

⋮

For k derivative

$$\xi^{(k)} = \partial^k \xi(\lambda, x) / \partial \lambda^k$$

$$= [(x+1)^\lambda \{\log(x+1)\}^k - k\xi^{(k-1)}] / \lambda$$

The first and second derivative of the (4) is given by

$$\xi(\lambda_-, x) = -\left\{\frac{(-x+1)^{\lambda-} - 1}{\lambda_-}\right\}$$

First derivative

$$\xi'(\lambda_-, x) = (-x+1)^{\lambda_- - 1}$$

Second derivative

$$\xi''(\lambda_-, x) = \lambda_- - 1(-x+1)^{\lambda_- - 2}(-1) |_{x=0}$$

$$= -\lambda_- + 1 \quad (7)$$

⋮

For k derivative

$$\xi^{(k)} = \partial^k \xi(\lambda, x) / \partial \lambda^k$$

$$= -[(-x+1)^{2-\lambda} \{-\log(-x+1)\}^k - k\xi^{(k-1)}] / (2-\lambda)$$

The first and the k derivative of the (3) is given by

$$\xi(\lambda_+, x) = \log(x+1)$$

First derivative

$$\xi'(\lambda_+, x) = \frac{1}{x+1}$$

Second derivative
$$\xi''(\lambda_+, x) = -\frac{1}{(x+1)^2}\Big|_{x=0} = -1 \tag{8}$$

$\vdots$ $\vdots$

For k derivative
$$\xi^{(k)} = \partial^k \xi(\lambda, x)/\partial\lambda^k = \{\log(x+1)\}^{k+1}/(k+1)$$

The first and the derivative of the (5) is given by

$$\xi(\lambda_+, x) = \log(-x+1)$$

First derivative
$$\xi'(\lambda_+, x) = \frac{1}{(-x+1)}$$

Second derivative
$$\xi''(\lambda_+, x) = -\frac{1}{(-x+1)^2}\Big|_{x=0} = -1 \tag{9}$$

$\vdots$ $\vdots$

For k derivative
$$\xi^{(k)} = \partial^k \xi(\lambda, x)/\partial\lambda^k = \{\log(-x+1)\}^{k+1}/(k+1)$$

The condition of the second derivative $\partial^2 \xi(\lambda_+, \lambda_-, x)/\partial x^2$ is continuous at $x = 0$, this imply the second derivative in (5) is equal to the second derivative in (6) as below

$$-\lambda_- + 1 = \lambda_+ - 1$$
$$\lambda_+ + \lambda_- = 2$$

This indicate that the transformation to be smooth and implies that $\lambda_+ + \lambda_- = 2$ and this induce $\lambda_+ = 2 - \lambda_-$. The sign of (8) and (9) indicate and also prove that the second derivative $\partial^2 \xi(\lambda_+, \lambda_-, x)/\partial x^2$ is continuous at $x = 0$.

Consequently, we define the power transformation, $\psi(.,.): R \times R \to R$ When the values of the X is greater or equal than zero, so we define

$$\xi(\lambda, x) = \rho_\lambda(X) = \begin{cases} \dfrac{(X+1)^\lambda - 1}{\lambda} & (x \geq 0, \lambda \neq 0), \\ \log(X+1) & (x \geq 0, \lambda = 0), \end{cases} \tag{10}$$

As the same case when the values of the X is smaller than zero, so we define

$$\xi(\lambda, x) = \rho_\lambda(X) = \begin{cases} -\left\{\dfrac{(-X+1)^{2-\lambda} - 1}{2-\lambda}\right\} & (x < 0, \lambda \neq 2), \\ \log(-X+1) & (x < 0, \lambda = 2), \end{cases} \tag{11}$$

From the second to fifth equation we add the constant 1 in parentheses to make the transformed value have the same sign as the original value, and allow us to prove Lemma 1 by working separately with the positive and negative domain.

Lemma 1. The transformation function $\xi(.,.)$ defines in (10) and (11) satisfy the following:

(i) $\xi(\lambda, x) \geq 0$ for $x \geq 0$ and $\xi(\lambda, x) < 0$ for $x < 0$;

(ii) $\xi(\lambda, x)$ is convex in x for $x > 1$ and concave in x for $\lambda < 1$;

(iii) $\xi(\lambda, x)$ is continuous function of (λ, x)

(iv) if $\xi^{(k)} = \partial^k \xi(\lambda, x) / \partial \lambda^k$ then for $k \geq 1$,

$$\xi^{(k)} = \begin{cases} [(x+1)^{\lambda}\{\log(x+1)\}^k - k\xi^{(k-1)}]/\lambda & (\lambda \neq 0, x \geq 0), \\ \{\log(x+1)\}^{k+1}/(k+1) & (\lambda = 0, x \geq 0), \\ -[(-x+1)^{2-\lambda}\{-\log(-x+1)\}^k - k\xi^{(k-1)}]/(2-\lambda) & (\lambda \neq 2, x < 0), \\ \{\log(-x+1)\}^{k+1}/(k+1) & (\lambda = 2, x < 0), \end{cases}$$

Note that $\xi^{(0)} \equiv \xi(\lambda, x)$

3. Conclusions

The original of Box-Cox formula is a simple method that can enable analysis of heteroscedastic and non-normal data sets so that the assumptions of the analysis of variance may be satisfied better, especially when other transformation procedures fail. As we know the Box-Cox formula is valid for the positive set data. A modified of the Box Cox transformation is proposed in this paper. It overcomes the truncation problem of the Box-Cox power transformation and also it generates a new family of transformation that can be applied to many other fields such as in economics, social sciences and many more.

4. Reference

Andrews, D. F., 1971. A note on the Selection of Data Transformation.Biometrika 58, 249-254.

Atkinson, A. C., 1982. Regression Diagnostics, Transformations and Constructed Variables with discussion. Journal of the Royal Stat Society Ser. B, 44, 1-36.

Bartlett, M. S., 1947. The Use of Transformations. Biometrics 3, 39-52.

Bickel, P. J. & Doksum, K. A., 1981. An Analysis of Transformations Revisited.
Journal of the American Statistician Association, 76, 269-311.

Box, G.E.P. & Cox, D.R., 1964. An Analysis of Transformations.
Journal of the Royal Stat. Society Series B, 26, 211-252.

Box, G.E.P. & Cox, D.R.,1982. An Analysis of Transformations Revisited, Rebutted. Journal of the American Statistician Association, 77, 209-210.

Chen, H., and Loh, W.Y.,1992. Bounds on ARES of test Based on Transformed DataAnnals of Statistics. Vol. 20, No.3, 1485-1500.

John, J.A., and Draper, N.R., 1980. An Alternative Family of Transformation.Applied Statistics 29, 190-197.

Hernandez, F. and Johnson, R.A,. 1980. The large-sample behavior of transformation toNormality. J. Amer. Statist. Assoc. 75, 855-861.

Hinckley, D.V., Runger, G,. 1984. Analysis of Transform Data. Journal of American Statistical Association. Volume 79. No. 386, 302-309.

Hinckley, D. V., 1975. On Power Transformations to Symmetry. Biometrical, 62, 101-111.

Hinckley, D. V. & Runger, G. 1984. Analysis of Transformed Data. Journal of the American Statistician Association, 79, 302-308.

Hoyle. M.H,. 1973. Transformation: An Introduction and a Bibliography. International Statistical Review 41, 203-23

Sakia R.M., 1992. The Box-Cox Transformation Technique : a review. The Statistician, 41, 169-178.

Tukey, J.W., 1957 On the Comparative Anatomy of Transformation. Annals of Mathematical Statistics. Volume 28, Issues 3, 602-632.

Lehmann, E.L., 1975. Nonparametric: Statistical Methods Based on Ranks. Holden-Day, San Francisco

Moore, P.G. and Tukey, J.W,. 1954. Answer to Query 112. Biometrics, 10, 562-568.

Miller, R.G,. 1986. Beyond ANOVA , Basic of Applied Statistics Wiley.