Statistika, Vol. 9 No. 1, 1 – 7
Mei 2009

Models for Transformation: A Global Optimization Transformation method with Some Extension from Box-Cox Transformation

WAN MUHAMAD AMIR W AHMAD[1], NYI NYI NAING [2]
AND TENGKU MOHAMAD ARIFF RAJA HUSSEIN[3]

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

ABSTRACT

The choice of a transformation has often been made in an ad hoc trial-and-error fashion but in this research paper we deals with the new solution of transformation which is covering all the real numbers. The assumption of normality that gained from the optimization method is much better improved compared to Box-Cox transformation through the statistical P value. The biggest values of the statistical P value (>0.05) reflect the goodness of the normality achievement. In order to obtain the efficiency status, we will illustrate the application of transformation method with the data that are getting from Hospital University Science Malaysia (HUSM).

Keywords: Box-Cox Formula, Alternative Formula, parameter λ and P- values

1. INTRODUCTION

When analyzing linear models, common assumptions for the response variable are independence, normality, and homoscedasticity. When one or more of these assumptions are violated a transformation of the response variable may be useful. Box-Cox (1964) considered the choice of a transformation among a parametric family of data transformation to yield an independence, normality, and homoscedasticity. They investigated two approaches to this problem and derived a likelihood function and a posterior distribution for the parameters of the transformation. In the year 1969, Draper and Cox have found approximation for the precision of the maximum likelihood estimate. By the year of 1967, Sir Fraser derived a different likelihood function which yields quite different inferences from those of Box-Cox in extreme cases where the number of parameters is close to the number of observations (Andrew, 1971). In the present paper, a method is proposed which has two possible advantages that's: (i) the amount of calculation is reduced if only one or a few transformation are to be tested; (ii) the precision with the transformation can be estimated is capable of theoretical calculation.

2. BACKGROUND AND METHOD

2.1 Box-Cox Transformation and Its Modification

The Box-Cox family of power transformation is widely used to achieve a normalizing transformation on a positive-valued response variable, Y. The family of power transformation is given by

$$\rho_{\lambda}(x)=\begin{cases}\dfrac{x^{\lambda}-1}{\lambda} & \lambda\neq 0\\ \ln x & \lambda=0\end{cases} \tag{1.1}$$

Box-Cox (1964) proposed maximum likelihood estimates of λ, β, and σ_e^2 in the linear model

$$W = X\beta + e \tag{1.2}$$

where $W=(W_1,W_2,\ldots W_n)'$, X is an $n\times p$ matrix of fixed regressor variables, β is a $p\times 1$ vector of unknown parameters and $e\sim N(0,\sigma_e^2 I)$. In this paper, we have considered the method of maximum likelihood for estimating λ, β, and σ_e^2. Some modifications have been studied and made to the formula given in the equation (1.1). The modulus term is being inserted to the formula in order to enhance the efficiency of transformation for the whole real numbers. The modification formula is given by

$$\psi(y,\lambda)=\begin{cases}\operatorname{sign}(y)\dfrac{|y|^{\lambda}-e^{\lambda}}{\lambda} & \lambda\neq 0\\ \operatorname{sign}(y)\ln(|y|)-1 & \lambda=0\end{cases} \tag{1.3}$$

The word sign in the equation above is referring to the sign of the original observations $y=(y_1,y_2,\ldots y_n)$ as such positive sign or negative sign. When we are taking limit to the equation (1.3) it's become $\lim_{\lambda\to 0}\operatorname{sign}(y)\dfrac{|y|^{\lambda}-e^{\lambda}}{\lambda}$. Solving them, it will lead to the $\operatorname{sign}(y)\ln(|y|)-1$.

$$
\begin{aligned}
\text{Let} \quad f(y) &= \operatorname{sign}(y)\frac{|y|^{\lambda}-e^{\lambda}}{\lambda} \\
&= \ln\left(\operatorname{sign}(y)\frac{|y|^{\lambda}-e^{\lambda}}{\lambda}\right) \\
&= \lim_{\lambda\to 0}\left[\ln\left(\operatorname{sign}(y)\frac{|y|^{\lambda}-e^{\lambda}}{\lambda}\right)\right] \\
&= \operatorname{sign}(y)\left(\lim_{\lambda\to 0}\left[\ln\left(\frac{|y|^{\lambda}-e^{\lambda}}{\lambda}\right)\right]\right) \\
&= \operatorname{sign}(y)\lim_{\lambda\to 0}\left(\ln|y|^{\lambda}-\ln(\lambda)-\left(\ln(e^{\lambda})-\ln(\lambda)\right)\right) \\
&= \operatorname{sign}(y)\lim_{\lambda\to 0}\left(\ln|y|^{\lambda}-\ln(e^{\lambda})\right) \\
&= \operatorname{sign}(y)\,\lambda\lim_{\lambda\to 0}\left(\ln|y|-\ln(e)\right) \\
&= \operatorname{sign}(y)\lambda\lim_{\lambda\to 0}\left(\left(\ln|y|-1\right)\right) \\
&= \frac{\operatorname{sign}(y)\lambda\lim_{\lambda\to 0}\left(\ln|y|-1\right)}{\lambda} \\
&= \operatorname{sign}(y)\left(\ln|y|-1\right)
\end{aligned}
$$

So, we can express the equation (1.3) as shown above. Let $\{\xi(y,\lambda)\}$ be a general family of transformation indexed by the transformation parameter λ. This could be family considered by modifying of Box-Cox (1964) and generally require that $\xi(y,\lambda)$ is monotone in y, otherwise, a model for $\xi(y,\lambda)$cannot produce a model for y by inverting the transformation. Throughout this paper it is assumed that $\xi(y,\lambda)$ has a second derivative that is continuous in λ. This allows us to make the maximum likelihood inference with respect to λ. We always see that an estimate of the parameter varies according to the selection of covariates and transformations

3. DETERMINATION THE POWER OF TRANSFORMATION (λ)

In this section we focus our attention on transforming a random sample from a parent distribution, with probability density function $f(\cdot)$, to near normality. Let $Y_1, Y_2, \ldots, Y_n$ be independent and identically distributed random variables and denote the transformed variables by $\xi(\lambda, Y_1), \ldots, \xi(\lambda, Y_n)$. We assumed that, for some λ, the transformed observation can be treated as normally distributed with some means μ and variance σ^2. Under this assumption, the log likelihood function is

$$
\begin{aligned}
L(\lambda,\mu,\sigma^2) = & -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\left(\frac{1}{n}\sum_{i=1}^{n}\left\{f(\lambda,y_i) - \frac{1}{n}\sum_{i=1}^{n} f(\lambda,y_i)\right\}^2\right) \\
& - \frac{1}{2\left(\frac{1}{n}\sum_{i=1}^{n}\left\{f(\lambda,y_i) - \frac{1}{n}\sum_{i=1}^{n} f(\lambda,y_i)\right\}^2\right)}\sum_{i=1}^{n}\left(f(\lambda,y_i) - \frac{1}{n}\sum_{i=1}^{n} f(\lambda,y_i)\right)^2 \\
& + (\lambda)\sum_{i=1}^{n}\operatorname{sign}(y)\log(|y_i|) - 1
\end{aligned}
\tag{1.3}
$$

We differentiate the equation in (1.3) respect to λ and gain the equation in (1.4) and the derivation for the equation in (1.4) is given by the in equation in (1.5).

$$
\begin{aligned}
\frac{d}{d\lambda}L(\lambda,\mu,\sigma^2) = & -\frac{1}{2} \\
& \left(n\left(\sum_{i=1}^{n}\left(\frac{2f(x_i,\lambda)\frac{\partial}{\partial\lambda}f(y_i,\lambda) - \frac{2\left(\frac{\partial}{\partial\lambda}f(y_i,\lambda)\right)\left(\sum_{i=1}^{n}f(y_i,\lambda)\right)}{n} - \frac{2f(y_i,\lambda)\left(\sum_{i=1}^{n}\left(\frac{\partial}{\partial\lambda}f(y_i,\lambda)\right)\right)}{n}}{}\right)\right)\right) \Bigg/ \\
& \left(\left(\sum_{i=1}^{n}\left(\frac{f(y_i,\lambda)^2 - 2\left(\frac{\partial}{\partial\lambda}f(y_i,\lambda)\right) - \frac{\left(\sum_{i=1}^{n}f(y_i,\lambda)\right)}{n}}{}\right)\right)\ln 10\right) \\
& + \sum_{i=1}^{n}\operatorname{sign}(y_i)\frac{\ln(|y|)}{\ln 10}
\end{aligned}
\tag{1.4}
$$

$$\frac{d^2L(\lambda,\mu,\sigma^2)}{d\lambda^2} =$$

$$-\frac{1}{2}\left(n\left(\sum_{j=1}^{n}\left(\begin{array}{l} 2\left(\frac{\partial}{\partial\lambda}f(x_i,\lambda)\right)^2 + 2f(x_i,\lambda)\left(\frac{\partial^2}{\partial\lambda^2}f(x_i,\lambda)\right) \\ -\frac{2\left(\frac{\partial^2}{\partial\lambda^2}f(x_i,\lambda)\right)\left(\sum_{i=1}^{n}f(x_i,\lambda)\right)}{n} \\ -\frac{4\left(\frac{\partial}{\partial\lambda}f(x_i,\lambda)\right)\left(\sum_{i=1}^{n}\left(\frac{\partial}{\partial\lambda}f(x_i,\lambda)\right)\right)}{n} \\ -\frac{2f(x_i,\lambda)\left(\sum_{i=1}^{n}\left(\frac{\partial^2}{\partial\lambda^2}f(x_i,\lambda)\right)\right)}{n} \end{array}\right)\right)\right) \Bigg/ \left(\left(\sum_{i=1}^{n}\left(f(x_i,\lambda)^2 - \frac{2f(x_i,\lambda)}{n}\right)\right)\ln(10)\right)$$

$$\frac{1}{2}\left(n\left(\sum_{i=1}^{n}\left(\begin{array}{l} 2f(x_i,\lambda)\left(\frac{\partial}{\partial\lambda}f(x_i,\lambda)\right) - \frac{2\left(\frac{\partial}{\partial\lambda}f(x_i,\lambda)\right)\left(\sum_{i=1}^{n}f(x_i,\lambda)\right)}{n} \\ -\frac{2f(x_i,\lambda)\left(\sum_{i=1}^{n}\left(\frac{\partial}{\partial\lambda}f(x_i,\lambda)\right)\right)}{n} \end{array}\right)^2\right)\right) \Bigg/$$

$$\left(\left(\sum_{i=1}^{n}\left(f(x_i,\lambda)^2 - \frac{2f(x_i,\lambda)\left(\sum_{i=1}^{n}f(x_i,\lambda)\right)}{n}\right)\right)^2\ln(10)\right) \qquad (1.5)$$

Holding the λ fixed, we initially maximize $L_n(\lambda,\mu,\sigma^2 \mid x)$, yielding

$$\hat{\mu}(\hat{\lambda}) = \frac{1}{n}\sum_{i=1}^{n} f(\lambda, x_i), \quad \hat{\sigma}^2(\hat{\lambda}) = \frac{1}{n}\sum_{i=1}^{n}\{f(\lambda, x_i) - \hat{\mu}(\lambda)\}^2$$

The estimate of $\hat{\lambda}$ is obtained by maximizing the profile loglikelihood function. We make use the Newton-Raphson method in order to calculate for the parameter $\hat{\lambda}$.

4. THE TEST OF SIGNIFICANCE FOR THE TRANSFORMED DATA

Consider $\Lambda = \{\lambda\}$, a parametric class of transformations

$$\lambda : y \to y^{(\lambda)} \ (\lambda \in \Lambda),$$

and suppose that for some λ the transformation response $y^{(\lambda)}$ may be describe by a linear model

$$y^{(\lambda)} = X\beta + \sigma e,$$

Where X is a $n \times p$ matrix of independent variables with rows x_i', β is a $p \times 1$ vector of unknown parameters, σ is an unknown scale parameter and e is a vector whose elements are

independent standard normal deviates. The statistics Anderson Darling (AD) and the statistics P value can be used in order to determine the significance of transformed data. Both of AD and AD* are needed for the calculation of statistical P value. The calculation formula for AD is given by

$$\mathrm{AD} = -\mathrm{n} - \frac{1}{\mathrm{n}}\sum_{\mathrm{i}=1}^{\mathrm{n}}(2\mathrm{n}-1)\left[\ln(\mathrm{z_i}) + \ln(1-\mathrm{z_{n-i+1}})\right]$$

and the modified formula of AD* is given by

$$\mathrm{AD}^* = \mathrm{AD}\left(1 + \frac{0.75}{\mathrm{N}} + \frac{2.25}{\mathrm{N}^2}\right)$$

Confidence Interval AD*	Calculation Formula for The p Value
$0.600 < AD^* < 13.($	$\mathrm{p} = \mathrm{e}^{\left(1.2937 - 5.709\,\mathrm{AD}^* + 0.0186(\mathrm{AD}^*)^2\right)}$
$0.340 < AD^* < 0.6($	$\mathrm{p} = \mathrm{e}^{\left(0.9177 - 4.279\,\mathrm{AD}^* - 1.38(\mathrm{AD}^*)^2\right)}$
$0.200 < AD^* < 0.3$	$\mathrm{p} = 1 - \mathrm{e}^{\left(-8.318 + 42.796\,\mathrm{AD}^* - 59.938(\mathrm{AD}^*)^2\right)}$
$AD^* < 0.200$	$\mathrm{p} = 1 - \mathrm{e}^{\left(-13.436 + 101.14\,\mathrm{AD}^* - 223.73(\mathrm{AD}^*)^2\right.}$

The quality of the transformed data can be measure through the statistical p value (>0.05). The biggest value reflects the goodness of the normality achievement. Same goes for the case of AD. The smallest of AD reflects the goodness of the normality achievement.

5. EXAMPLE

We will illustrate the application of the transformation by using the data that gain from Hospital University Sains Malaysia (HUSM). The reading of difference weight before and after treatment is measured in order to know the effectiveness of the medicine taken. 15 patients were involved in this study and the data were given as 7.1, -9.4, 1.1, 2, 0.7, 3.9, 3.5, 8.1, 2.8, 5.6, 7.0, 6.0, 9.3, 3.5 and -2.0.

For the original data, statistics Anderson Darling is given by $AD = 0.49$, the skewness value is given by -1.42799 $(p = 0.182)$, with the mean 3.2800 and variance 21.6503. When the new transformation is applied the parameter estimate for (λ, μ, σ^2) are

$$\begin{pmatrix} \hat{\lambda} \\ \hat{\mu} \\ \hat{\sigma}^2 \end{pmatrix} = \begin{pmatrix} 1.452 \\ 2.708 \\ 50.056 \end{pmatrix}$$

For the transformed data, the calculated AD and the skewness statistics is given by 0.39 and -0.67853 (with $\mathrm{p} = 0.333$). The statistical value showing that the normality of transformed data is much improved.

Figure 1 shows that the normal approximation is much improved since the transform pull down the right tail and pushed out the left tail.

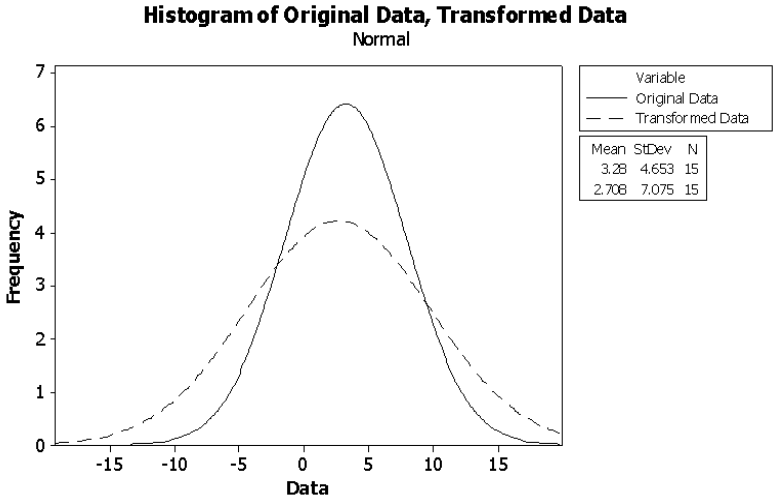

Figure 1. The Differences of Normality Curve Between Original and Transformed Data

6. CONCLUSION

A new dimension of transformation method is being exposed in this research paper. In order to gain the optimize value of transformed data. This alternatives method is introduced and it is simple enough to be used by practitioner, computed with the output of standard statistics packages, and covered in an applied statistics course.

Acknowledgements

The author is very grateful to Professor Syed Hatim Noor and Professor Tengku Mohamad Ariff for very helpful comments which led to the present improved version of the article.

REFERENCES

[1]. Andrews, D. F. (1971). A note on the Selection of Data Transformation. *Biometrika* 58: 249-254.
[2]. Box, G.E.P. & Cox, D.R. (1964). An Analysis of Transformations. *Journal of the Royal Stat. Society Series B*, 26: 211-252.
[3]. Box, G.E.P. & Cox, D.R. (1982). An Analysis of Transformations Revisited, Rebutted. *Journal of the American Statistician Association*, 77, 209-210.
[4]. Sakia R.M. (1992). The Box-Cox Transformation Technique: A review. *The Statistician*, 41:169-178.
[5]. Yeo, I. K. (2005). Variable Selection and Transformation in Linear Regression Models. Elsevier 72, 219-226.