*Academic Journal of Nawroz University (AJNU), Vol.12, No.4, 2023*

https://doi.org/10.25007/ajnu.v12n4a1902

# Transforming to Normality in Regression Analysis with Exponentially Residuals

Azad Adil Shareef [1], Haithem Taha Mohammed Ali [2,3]

[1] Department of Statistics, University of Duhok, Kurdistan Region, Iraq
[2] Department of Economic Sciences, University of Zakho, Kurdistan Region, Iraq
[3] Department of Economics, Nawroz University, Kurdistan Region, Iraq

## ABSTRACT

In this article, a simulated study is introduced, focusing on the use of power transformation to estimate a nonlinear regression model in the presence of residuals following an exponential distribution. Four criteria were employed to estimate the power parameter: the p-value of Shapiro-Wilk test statistics for both the transformed and back-transformed data's normality, maximum likelihood estimation, and coefficient of determination. The findings of the study indicate that while it is possible to identify a range of viable solutions to select the optimal power parameter, finding a single optimal value that satisfies all estimation and decision methods is not feasible.

**KEY WORDS**: Multiple linear regression, Exponential residuals, Box-Cox transformation, Shapiro Wilk Test.

## 1. INTRODUCTION

In linear regression, one of the key assumptions is that the residuals, which are the differences between the observed values and the predicted values, should follow a normal distribution. However, the original observed values and the predicted values themselves are not required to be normally distributed. The normality assumption for residuals is important for various statistical inferences and assessing the reliability of the model, but it does not apply to the raw data or the predictions made by the model. In practice, deviations from normality in the observed values or predictions might not be problematic as long as the normality assumption holds for the residuals. Nonetheless, it is good practice to check for normality and consider alternative approaches if the assumption is significantly violated.

Exponential regression (ER) is a regression analysis technique that involves fitting an exponential function to data points. This approach allows us to estimate the function's parameters, enabling prediction and drawing conclusions based on the fitted curve. Its application proves particularly useful for modeling and analyzing data demonstrating exponential growth or decay patterns. Many real-world scenarios exhibit such exponential behavior, including population growth, the decay of radioactive substances, the spread of infectious diseases, and asset depreciation over time. By providing a mathematical framework, ER facilitates a deeper understanding and quantification of these exponential relationships.

ER model assumes that the relationship between the dependent variable vector $\boldsymbol{Y}$ with size $(n \times 1)$ and the explanatory variables $X$ matrix with size $(n \times p)$ follows an exponential function of the form:

$$\boldsymbol{Y} = \beta_0 \, exp\{\boldsymbol{X\beta}\} + \boldsymbol{U} \qquad (1)$$

Where $\beta_0$ represents the initial value of the dependent variable at $\boldsymbol{X} = \boldsymbol{0}$, $\boldsymbol{\beta}$ is the vector of $p$ parameters with size $(p \times 1)$ of exponential growth

or decay and $\boldsymbol{U}$ is the residuals vector with size $(n \times 1)$.

The objective of ER is to determine the optimal estimates for the coefficients $\beta_0$ and $\boldsymbol{\beta}$ that align most effectively with the given dataset. This is accomplished through the minimization of the following sum of squared variances between the real data points and the predicted values generated by the exponential function,

$$SSE = (\boldsymbol{Y} - \beta_0 \exp\{\boldsymbol{X\beta}\})^T (\boldsymbol{Y} - \beta_0 \, exp\{\boldsymbol{X\beta}\}) \quad (2)$$

Through the process of fitting the exponential curve to the data, we derive estimates for $\beta_0$ and $\boldsymbol{\beta}$ vector, representing the overall trend and growth or decay rate. These estimates empower us to make predictions beyond the observed data range and gain valuable insights into the behavior of the studied system or phenomenon.

ER finds extensive application in diverse fields, such as biology, economics, finance, epidemiology, and environmental sciences. This widely-used technique offers a potent tool for analyzing and comprehending exponential relationships, equipping researchers, analysts, and decision-makers with the means to make informed predictions and sound decisions based on the observed data [1].

It is crucial to recognize that ER is well-suited for datasets displaying exponential behavior, but it may not be suitable for other types of relationships. It is vital to carefully examine the data's nature and explore alternative regression techniques if the relationship does not exhibit exponential characteristics.

In this article, we focus on situations where the residuals appear to have an exponential pattern. According to statistical theory, this implies that the dependent variable in multiple regression follows an exponential distribution. Consequently, the model belongs to the class of models equivalent to Eq. 1. In linear regression analysis, residuals play a crucial role as a diagnostic tool to evaluate the model's adequacy and the underlying assumptions of the analysis. When the residuals display an exponential pattern, it could indicate potential systematic issues in the model, such as heteroscedasticity or a non-linear relationship between the predictors and the response variable [2].

In multiple linear regression (MLR), $\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{U}$, if we assume $\boldsymbol{b}$ is the vector of $\boldsymbol{\beta}$ estimate and the residuals vector $\boldsymbol{U} = (\boldsymbol{Y} - \boldsymbol{Xb}) > \boldsymbol{0}$ follow the exponential distribution of the form $f_{\boldsymbol{U}}(u_i) = \theta \exp(-\theta u_i)$, then the dependent variable $\boldsymbol{Y}$ follow the following probability density function (PDF),

$$f_{\boldsymbol{Y}}(y_i) = \theta \exp\{-\theta \, (y_i - \boldsymbol{b})\} \quad , y_i > \boldsymbol{b} \quad (3)$$

where, $i = 1, 2, \dots, n$. In practical applications, it is crucial to understand that the distributional assumption made for the residuals does not necessarily imply a direct impact on the dependent variable. The distributional assumption for the dependent variable is distinct and may follow a different distribution or exhibit different characteristics altogether.

Little. R. conducted a simulation study in 1979 to explore maximum likelihood inference for coefficients in a multiple regression model. This approach was contrasted with the least squares method using mean squared error [3]. In 1998, Rawlings, Pantula, and Dickey highlighted the critical conditions for statistical inference in the MLR model: normality and variance constancy of the estimated model residuals [4]. Consequently, data transformation techniques, especially those belonging to the power transformation family, have been widely used to improve the utility of statistical modeling and achieve a better fit.

One common strategy to address non-normality in residuals is through transformations. This aims to approximate normal distribution of residuals, satisfying the normality assumption. When dealing with residuals exhibiting an exponential pattern, various transformations can be applied to achieve normality.

The Box-Cox transformation (BCT) approach is designed to meet the conditions of modeling in MLR by using parametric power transformation [5]. Yeo and Johnson introduced a new family of distributions in 2000, possessing many desirable properties of BCT and usable without restrictions for cases with positive and negative variable values [6]. In 2004, Nishidate and Mishina proposed a technique for analyzing diffuse reflectance spectra from skin tissue using multiple regression analysis combined with a Monte Carlo simulation. This approach yielded regression coefficients, utilizing the absorbance spectrum as the response variable and the extinction coefficients of melanin, oxygenated hemoglobin, and deoxygenated hemoglobin as predictor variables [7].

Various methodologies have been proposed and studied in the field of statistical analysis. One such approach is the multivariate regression modeling with functional data analysis, as presented by Matsui et al. (2008). Their study demonstrates how Gaussian basis functions and regularization techniques can establish the relationship between multiple scalar responses and functional predictors. To validate the efficacy of their proposed method, Monte Carlo

simulations and real data (Spectrometric data) were employed [8]. In 2011, Hossain conducted an analytical review focusing on the significant role of BCT technique in various statistical domains such as estimation, testing, inference, and model selection.[9]. Another research in 2014 by Al-Yousef and Abduahad introduced a method for describing Bayesian conditional expectation Probability distribution (gamma - gamma) through a nonlinear regression model using power transformation. [10]. In 2016, Fischer put forth the idea of utilizing tree growth models and the logarithmic transformation to address left-sided variable transformation in order to meet the assumptions of regression analysis. [11]. Recently, in 2021, Atkinson, Riani, and Corbellini worked on the BCT of non-negative responses in linear regression models. Their extensions covered the transformation of both sides of the model and the Yeo-Johnson transformation for observations that can be positive or negative.[12]. In the year 2022, Al-Safar and Mohammed Ali made use of power transformation to enhance nonlinear models of the response surfaces methodology. This approach aimed to improve the accuracy and reliability of these models.[13].

The article emphasizes the widespread utility of the BCT, a versatile transformation function applicable to various regression models, including ER. This article delves into the application of BCT within the context of ER and its potential to normalize distributions, address normality assumptions, stabilize variance, and enhance the linearity between variables.

The main goal of this study is to utilize BCT in ER when confronted with residuals following an exponential distribution. It's important to note that when residuals exhibit an exponential distribution, the response variable must also adhere to the same distribution.

The article's organization is as follows: The second section covers material and methods, incorporating theoretical aspects related to ER. The third section introduces the proposed algorithm and its application in developing a nonlinear regression model using BCT. Subsequently, the fourth section presents and discusses the results obtained. Finally, the fifth section provides the conclusions drawn from the study's outcomes.

## 2. MATERIAL AND METHOD

In ER, the logarithm transformation is frequently employed to convert the exponential relationship between variables into a linear one. This transformation enables the use of linear regression techniques to fit a line to the transformed data, involving the estimation of coefficients like intercepts and slopes. By fitting the linear regression model on the transformed data, we can interpret the results in terms of the original variables. The application of logarithmic transformations often leads to a better linear fit and enhances the interpretability of the relationship between variables. However, it's essential to consider the appropriateness of the logarithm transformation based on specific data characteristics and the suitability of the linear model for the transformed data. Therefore, while the logarithmic transformation can transform the ER function into a linear one, its efficiency in representing the data may vary and requires careful assessment.

Based on the above, a more general transformation function may be used instead of the specific logarithm transformation. This allows for more flexibility in capturing the underlying relationship between variables in ER.

For the response vector $\boldsymbol{Y}$ in Eq. 1, BCT is defined by the formula:

$$\boldsymbol{Y}^{(\lambda)} = \varphi(y) = \begin{cases} y^{\lambda} - 1/\lambda & if \;\; \lambda \neq 0 \\ Ln\,(y) & if \;\; \lambda = 0 \end{cases} \tag{4}$$

When the BCT is applied with $\lambda = 1$, it corresponds to analyzing the data in its original scale without any transformation using the following MLR model,

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{U} \tag{5}$$

In other words, no transformation is applied, and the response vector is used as it is in the analysis. The logarithmic transformation can be seen as a specific instance of BCT when $\lambda = 0$. This means, the data analysis will be according to the form, $ln\,\boldsymbol{Y} = f(\boldsymbol{XB}) + \boldsymbol{U}$. The back transformation of BCT of response vector involves transforming the transformed vector back to its original scale and defined as the following nonlinear model:

$$\boldsymbol{Y} = \begin{cases} \left(\lambda\, \boldsymbol{X\widehat{\beta}} + 1\right)^{1/\lambda} & if\ \lambda \neq 0 \\ exp\ (\boldsymbol{X\widehat{\beta}}) & if\ \lambda = 0 \end{cases} \tag{6}$$

A widely employed conventional approach for estimating the model and the power parameter $\lambda$ is through the method of maximum likelihood estimation (MLE). If we suppose that $\boldsymbol{Y}^{(\lambda)} = \boldsymbol{X\beta} + \boldsymbol{U}$ represents the linear model of the transformed data so that $\boldsymbol{Y}^{(\lambda)} \sim N(\boldsymbol{X\beta}, \sigma_{\varepsilon}^2)$, the Likelihood to estimate the power Parameter and $\boldsymbol{\beta}$ vector $L_Y(y; \lambda, \boldsymbol{\beta}, \sigma_u^2) =$

$\prod_{i=1}^{n} f_Y(\varphi(y);\ \lambda, \boldsymbol{X\beta}, \sigma_u^2)|d\varphi(y)/dy|$ according to the following joint PDF,

$$L_Y(y;\lambda,\boldsymbol{\beta},\sigma_u^2) = (\,2\pi\sigma_\varepsilon^2)^{-\mathrm{n}/2} \times exp\left\{-\sum_{i=1}^{\mathrm{n}} (\varphi(\mathrm{y}) - \mathbf{X\beta})^{\mathrm{T}}(\varphi(\mathrm{y}) - \mathbf{X\beta})/2\sigma_\varepsilon^2\right\} \times \prod_{i=1}^{n}|d\varphi(\mathrm{y})/dy| \quad (7)$$

Then, the log likelihood is as the following form:

$$log_e\, L(y;\lambda,\boldsymbol{\beta},\sigma_\varepsilon^2) = (-\mathrm{n}/2)\, log_e\, 2\,\pi - (\mathrm{n}/2)\, log\, \sigma_\varepsilon^2 \ - \sum SSE/2\sigma_\varepsilon^2 + \sum_{i=1}^{n} log_e|d\varphi(\mathrm{y})/dy| \quad (8)$$

where:

$SSE = (\varphi(\mathrm{y}) - \mathbf{X\beta})^{\mathrm{T}}(\varphi(\mathrm{y}) - \mathbf{X\beta})$

Thus, the optimal $\lambda$ is the value that maximize Eq. 8.

## 3. COMPUTATIONAL ALGORITHM AND APPLICATION

In this section, the authors proposed an application algorithm of the using of BCT model and parametric estimation to develop a nonlinear multiple regression model when the residuals follow exponential distribution and they used different criteria to estimate the parameter. The choosing of optimum power parameter $\lambda$ in this algorithm is based on the four following different criteria; The first and second criteria are the coefficient of determination and MLE that can be used for original random variable $Y$. The third criteria are applied to select highest value of p-value of Shapiro-Wilk test for the residual normality resulting from the estimated nonlinear regression model of the original data vector. The last criterion is to choose of the highest value of p-value of Shapiro-Wilk test for the residual normality resulting from back transformed data. Therefore, the proposed application algorithm of the using of BCT model and parametric estimation to develop a nonlinear multiple regression model were shown in the algorithm steps. We conducted a simulated study in which data sets from normal distribution and the residuals followed exponential distribution with the parameters that could be used to estimate the effectiveness of the highest value MLE and p-value of Shapiro Wilk test which was based on the technique that could be mentioned the residuals for simulated model distributed as an exponential. Consider fitting a regression model where the response variable is a linear model with combination of 3 explanatory variables plus random residual. Consider an example in which we can generate random values for the explanatory variables and do not need to use real $\boldsymbol{X}$ values. R program is used to generated three samples with different sample size. The response variable and independent variables are simulated via the following steps and we implemented the following algorithm in R.

Step 1: Consider the following assumptions of MLR models to be generated: (a) Three explanatory variables vectors $\boldsymbol{X}_1, \boldsymbol{X}_2,$ and $\boldsymbol{X}_3$**.** (b) Three assumed sample sizes, $n = 15, 30$ and $50$. (c) The residuals distributed as exponential distribution with the PDF, $f_U(u) = \theta\, \exp(-\theta u)\,, U > 0$, fix $\theta = 2, 4, 8$ . (d) $\boldsymbol{\beta}$ **t**he vector of parameters, intercept and slopes is bounded by the rang $(0, 1)$, fix $\boldsymbol{\beta}' = [0.5 \quad 0.7 \quad 0.8 \quad 0.3]$. (e) The explanatory variables distributed as normal distribution.

Step 2: Generate the data sets of that will be used, each of which includes an independent variable vector $\boldsymbol{Y}$ and three explanatory variables based on the assumptions of step 1.

Step 3: Choose a set of candidate values for the power parameter of BCT. Fix $\lambda \in \Lambda$, where $\Lambda = \{-a, -(a - 0.1), \dots, 0\,, \dots, (a - 0.1)\,, a\}$. $a$ is an integer and can be chosen from which we can obtain a convex curve for MLE. For all generated data sets we perform the following steps 4 to 9.

Step 4: Transform original response variable $Y$ to $\psi(y)$ using BCT according to Equation (4).

Step 5: Estimate MLR model of the transformed data vector $\psi(\mathrm{y})/X_1\,, X_2, X_3,$ according to Equation (5).

Step 6: Estimate log likelihood function $L_{max}(\lambda, y)$ according to Equation (8). Calculate COD, p-value of SWT statistics to test the residual vector normality

Step 7: Calculate the P-value of Shapiro-Wilk test of the residual vector normality of the estimated nonlinear model of the original data vector.

Step 8: Calculate the P-value of Shapiro-Wilk test of the residual vector normality of the estimated nonlinear model of the back transformed data.

Step 9: Repeat all the steps from 3 to 8 for all values of $\lambda$ in $\Lambda$.

Step 10: For all the steps from 1 to 9 repeat the calculations 25 replicate.

## 4. RESULTS AND DISCUSSION

The computational algorithm was implemented using R software. The results in Table 1 demonstrate the estimation of the power parameter based on four criteria, utilizing the proposed algorithm for three different sample sizes. For a sample size of 15, the p-value of the Shapiro-Wilk test statistics for the residual's normality of the estimated regression on the transformed response vector falls within the range $(0.29, 0.99)$, indicating normal distribution. The optimal power parameter is $\lambda \epsilon\, (-1, 3)$. Similarly, for the back-transformed data, the p-value lies within the range $(0.18, 0.99)$, and the optimal power parameter is $\lambda \epsilon\, (-0.9, 3)$. COD for the

transformed response vector varies from 0.12 to 0.97, with the optimal power parameter $\lambda \epsilon (-3, 3)$. Additionally, the MLE values range from -27.0 to 1.0, with the optimal power parameter $\lambda \epsilon (-0.1, 0.9)$. As the sample size increases to 30 and 50, similar patterns emerge, and the optimal power parameter values are within the specified ranges.

Table 2 presents the results when $\varepsilon_i \sim exp(\theta = 4)$ for three different sample sizes. The optimal power parameter for the p-value of the Shapiro-Wilk test statistics for both the transformed response vector and back-transformed data is $\lambda \epsilon (-2.6, 3)$. The maximum COD value for the power parameter is $\lambda \epsilon (-3, 3)$, and the optimal MLE power parameter is $\lambda \epsilon (0.2, 1.1)$.

Similarly, Table 3 illustrates the outcomes for $\varepsilon_i \sim exp(\theta = 8)$ for three different sample sizes. The optimal power parameter for the p-value of the Shapiro-Wilk test statistics for both the transformed response vector and back-transformed data is $\lambda \epsilon (-1, 3)$. The optimal COD and MLE power parameters are $\lambda \epsilon (-3, 3)$ and $\lambda \epsilon (-0.1, 0.9)$, respectively.

The results indicate that as the sample size increases, the p-values for the Shapiro-Wilk test statistics tend to become smaller, indicating a better fit to normality. In conclusion, for simulated data with different sample sizes, MLE provides the best criteria, with the optimal power parameter range being $\lambda \epsilon (-0.1, 1.1)$, which shows consistent values across different scenarios.

**Table 1**: Simulating data from different sample size when $\epsilon_i \sim exp\ (\theta = 2)$

| Criteria | | Sample size (n) | | |
|---|---|---|---|---|
| | | 15 | 30 | 50 |
| **MLE** | values | (-27.0, 1.0) | (-49.0, -20.9) | (-94.7, -67.9) |
| | λ | (-0.1, 0.9) | (0.2, 0.9) | (0.4, 0.8) |
| **COD** | values | (0.12, 0.97) | (0.14, 0.92) | 0.09, 0.67) |
| | λ | (-3, 3) | (0.2, 3) | (-0.2, 3) |
| **P-v. BT** | values | (0.18, 0.99) | (0.01, 0.98) | (0.02, 0.99) |
| | λ | (-0.9, 3) | (-2.5, 2.9) | (-2.4, 3) |
| **P-v. DT** | values | (0.29, 0.99) | (0.0, 0.99) | (0.0, 0.91) |
| | λ | (-1, 3) | (-0.9, 0.8) | (0.0, 3) |

**Table 2**: Simulating data from different sample size when $\epsilon_i \sim \exp(\theta = 4)$

| Criteria | | Sample size (n) | | |
|---|---|---|---|---|
| | | 15 | 30 | 50 |
| **MLE** | values | (-19.8, 22.4) | (-51.8, -22.8) | (-93.7, -49.6) |
| | λ | (0.2, 1.1) | (0.5, 0.9) | (0.3, 0.8) |
| **COD** | values | (0.38, 0.99) | (0.30, 0.87) | (0.24, 0.74) |
| | λ | (-3, 3) | (0, 3) | (0.7, 2.4) |
| **P-v. BT** | values | (0.66, 0.99) | (0.13, 0.99) | (0.06, 0.98) |
| | λ | (-2.6, 3) | (-1.8, 3) | (-1.9, 3) |
| **P-v. DT** | values | (0.34, 0.99) | (0.05, 0.99) | (0.0, 0.96) |
| | λ | (-0.8, 3) | (0.3, 2.4) | (-0.2, 1) |

Table 3: Simulating data from different sample size when $\epsilon_i \sim \exp\ (\theta = 8)$

| Criteria | | Sample size (n) | | |
|---|---|---|---|---|
| | | 15 | 30 | 50 |
| **MLE** | values | (-26.9, 0.99) | (-47.9, -20.8) | (-94.7, -67.9) |
| | λ | (-0.1, 0.9) | (0.2, 0.9) | (0.4, 0.8) |
| **COD** | values | (0.12, 0.97) | (0.14, 0.92) | (0.09, 0.67) |
| | λ | (-3, 3) | (0.2, 3) | (-0.2, 3) |
| **P-v. BT** | values | (0.18, 0.99) | (0.01, 0.98) | (0.02, 0.99) |
| | λ | (-0.9, 3) | (-2.5, 2.9) | (-2.4, 3) |
| **P-v. DT** | values | (0.29, 0.99) | (0.0, 0.99) | (0.0, 0.91) |
| | λ | (-1, 3) | (-0.9, 0.8) | (0.0, 3) |

## 5. CONCLUSION

The combined p-value approximation proves to be a superior method for assessing normality, as it considers the agreement between transformed and p-value estimates. Through our analysis of normal distribution and data, we deduce that a single value of the p-value cannot guarantee normality. However, the likelihood of obtaining significantly larger p-values for multiple values appears implausible.

Importantly, the optimal power parameters derived from the transformation models using BCT demonstrate significant effectiveness. Various methods exist for selecting the best power parameter. The first approach involves using well-known estimation methods like the MLE method. The second method employs efficiency criteria from regression modeling, such as the COD and p-value, as decision rules for power parameter estimation, as demonstrated in this study.

In conclusion, for simulated data with different sample sizes, the MLE method proves to be the best criterion for determining the optimal power parameter. The range of optimal power parameters is $\lambda \epsilon (0.0, 1.1)$, and these values closely align with each other.

## REFERENCES

[1] J. Beirlant, G. Dierckx, Y., Goegebeur, and G. Matthys, Tail index estimation and an exponential regression model. Extremes, 2,

pp.177-200, 1999.

[2] Montgomery, D. C., Peck, E. A., & Vining, G. G. Introduction to Linear Regression Analysis (5th ed.). Wiley, 2012.

[3] R. J. A. Little, "Maximum likelihood inference for multiple regression with missing values: A simulation study," *J. R. Stat. Soc. Ser. B*, vol. 41, no. 1, pp. 76–87, 1979.

[4] J. O. Rawlings, S. G. Pantula, and D. A. Dickey, *Applied regression analysis: a research tool*. Springer, 1998.

[5] G. Box, "EP and Cox, D. R.'," *An Anal. Transform. J. R. Stat. Soc. Ser. B*, vol. 26, pp. 211–252, 1964.

[6] I. N. K. Yeo and R. A. Johnson, "A new family of power transformations to improve normality or symmetry," *Biometrika*, vol. 87, no. 4, pp. 954–959, 2000, doi: 10.1093/biomet/87.4.954.

[7] I. Nishidate, Y. Aizu, and H. Mishina, "Estimation of melanin and hemoglobin in skin tissue using multiple regression analysis aided by Monte Carlo simulation," *J. Biomed. Opt.*, vol. 9, no. 4, pp. 700–710, 2004.

[8] H. Matsui, Y. Araki, and S. Konishi, "Multivariate regression modeling for functional data," *J. Data Sci.*, vol. 6, no. 3, pp. 313–331, 2008.

[9] Hossain M. Z., "The use of Box-Cox transformation technique in economic and statistical analyses," *J. Emerg. Trends Econ. Manag. Sci.*, vol. 2, no. 1, pp. 32–39, 2011.

[10] Alyousif H. T. and and Abduahad F.N., "Develop a Nonlinear Model for the Conditional Expectation of the Bayesian Probability Distribution (Gamma – Gamma)," *J. Al-Nahrain Univ.*, vol. 17, no. 2, pp. 205–212, 2014.

[11] C. Fischer, "Comparing the logarithmic transformation and the Box-Cox transformation for individual tree basal area increment models," *For. Sci.*, vol. 62, no. 3, pp. 297–306, 2016, doi: 10.5849/forsci.15-135.

[12] Atkinson A.C. and R.M. and C.A., *The Box–Cox Transformation: Review and Extensions*, vol. 36, no. 2. 2021, pp. 239–255. doi: 10.1214/20-STS778.

[13] A. Al-Saffar and H. T. M. Ali, "Using Power Transformations in Response Surface Methodology," in *2022 International Conference on Computer Science and Software Engineering (CSASE)*, 2022, pp. 374–379.

.