Transformations

Tiago M. Magalhães

Department of Statistics - ICE-UFJF

Juiz de Fora, November 06, 2026

Outline

1. Introduction
2. Variance-stabilizing transformations
3. Box-Cox transformation
4. Generalized least squares
5. Application
6. References

Outline

1. Introduction
2. Variance-stabilizing transformations
3. Box-Cox transformation
4. Generalized least squares
5. Application
6. References

Linear regression model

Suppose that $Y_1, Y_2, \ldots, Y_n$ satisfy

$$Y_\ell = \boldsymbol{x}_\ell^\top \beta + \varepsilon_\ell, \tag{1}$$

Linear regression model

Suppose that $Y_1, Y_2, \ldots, Y_n$ satisfy

$$Y_\ell = \boldsymbol{x}_\ell^\top \boldsymbol{\beta} + \varepsilon_\ell, \tag{1}$$

where $\boldsymbol{x}_\ell = (x_{\ell 1}, x_{\ell 2}, \ldots, x_{\ell p})^\top$ is known, $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_p)^\top$ is a vector of unknown parameters to be estimated, $\varepsilon_\ell \sim \mathcal{D}(0, \sigma^2)$, $\ell = 1, 2, \ldots, n$, $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ are independent random variables with mean zero and common variance σ^2, also unknown, to be estimated.

Linear regression model

Suppose that $Y_1, Y_2, \ldots, Y_n$ satisfy

$$Y_\ell = \boldsymbol{x}_\ell^\top \boldsymbol{\beta} + \varepsilon_\ell, \tag{1}$$

where $\boldsymbol{x}_\ell = (x_{\ell 1}, x_{\ell 2}, \ldots, x_{\ell p})^\top$ is known, $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_p)^\top$ is a vector of unknown parameters to be estimated, $\varepsilon_\ell \sim \mathcal{D}(0, \sigma^2)$, $\ell = 1, 2, \ldots, n$, $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ are independent random variables with mean zero and common variance σ^2, also unknown, to be estimated.

Matrix form

Equation (1), which we defined as the **linear regression model** (LRM), can be written in matrix form as follows:

$$\boldsymbol{Y} = \boldsymbol{X}\beta + \varepsilon, \tag{2}$$

Matrix form

Equation (1), which we defined as the **linear regression model** (LRM), can be written in matrix form as follows:

$$\boldsymbol{Y} = \boldsymbol{X}\beta + \boldsymbol{\varepsilon}, \tag{2}$$

where $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_n)^\top$, $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)^\top$, $\boldsymbol{\varepsilon} \sim \mathcal{D}_n(\mathbf{0}, \sigma^2 \boldsymbol{I}_n)$, $\mathbf{0}$ is the zero vector of dimension n, $\boldsymbol{I}_n$ is the identity matrix of order n, and $\boldsymbol{X} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n)^\top$ is the design matrix.

Matrix form

Equation (1), which we defined as the **linear regression model** (LRM), can be written in matrix form as follows:

$$\boldsymbol{Y} = \boldsymbol{X}\beta + \boldsymbol{\varepsilon}, \tag{2}$$

where $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_n)^\top$, $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)^\top$, $\boldsymbol{\varepsilon} \sim \mathcal{D}_n(\mathbf{0}, \sigma^2 \boldsymbol{I}_n)$, $\mathbf{0}$ is the zero vector of dimension n, $\boldsymbol{I}_n$ is the identity matrix of order n, and $\boldsymbol{X} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n)^\top$ is the design matrix.

Assumptions

In summary, the **assumptions** of a linear regression model are:

- The relationship between the response and predictor variables is linear;

Assumptions

In summary, the **assumptions** of a linear regression model are:

- The relationship between the response and predictor variables is linear;
- The errors

Assumptions

In summary, the **assumptions** of a linear regression model are:

- The relationship between the response and predictor variables is linear;
- The errors
  - have mean zero;

Assumptions

In summary, the **assumptions** of a linear regression model are:

- The relationship between the response and predictor variables is linear;
- The errors
  - have mean zero;
  - have constant variance;

Assumptions

In summary, the **assumptions** of a linear regression model are:

- The relationship between the response and predictor variables is linear;
- The errors
  - have mean zero;
  - have constant variance;
  - are uncorrelated.

Assumptions

In summary, the **assumptions** of a linear regression model are:

- The relationship between the response and predictor variables is linear;
- The errors
  - have mean zero;
  - have constant variance;
  - are uncorrelated.

Outline

1 Introduction

2 Variance-stabilizing transformations

3 Box-Cox transformation

4 Generalized least squares

5 Application

6 References

Variance-stabilizing transformations

The constant variance assumption is often violated when the variance is related to the mean.

A model with non-constant variance often yields regression coefficient estimates with inflated standard errors.

Variance-stabilizing transformations

The constant variance assumption is often violated when the variance is related to the mean.

A model with non-constant variance often yields regression coefficient estimates with inflated standard errors.

Variance-stabilizing transformations

Transformations on the response variable can eliminate this problem.

The effect of the transformation depends on the amount of curvature it induces.

Variance-stabilizing transformations

Transformations on the response variable can eliminate this problem.

The effect of the transformation depends on the amount of curvature it induces.

Variance-stabilizing transformations

Table 1: Useful transformations.

Relationship between σ^2 and $\mathbb{E}(Y)$	Transformation	Remarks
$\sigma^2 \propto$ constant	$Y^* = Y$	No transformation
$\sigma^2 \propto \mathbb{E}(Y)$	$Y^* = \sqrt{Y}$	Poisson data
$\sigma^2 \propto \mathbb{E}(Y)[1 - \mathbb{E}(Y)]$	$Y^* = \text{arcsin}(\sqrt{Y})$	Binomial data
$\sigma^2 \propto [\mathbb{E}(Y)]^2$	$Y^* = \log(Y)$	
$\sigma^2 \propto [\mathbb{E}(Y)]^3$	$Y^* = Y^{-1/2}$	
$\sigma^2 \propto [\mathbb{E}(Y)]^4$	$Y^* = Y^{-1}$	

Outline

1 Introduction

2 Variance-stabilizing transformations

3 Box-Cox transformation

4 Generalized least squares

5 Application

6 References

Box-Cox transformation

A useful class of power transformations to correct non-normality and/or non-constant variance.

Proposed by Box and Cox (1964), it is a class of transformations given by

$$Y^* = Y(\lambda),$$

where λ is an additional parameter to be estimated.

Box-Cox transformation

A useful class of power transformations to correct non-normality and/or non-constant variance.

Proposed by Box and Cox (1964), it is a class of transformations given by

$$Y^* = Y(\lambda),$$

where λ is an additional parameter to be estimated.

Box-Cox transformation

The procedure is defined as follows:

$$Y^* = \begin{cases} \dfrac{Y^\lambda - 1}{\lambda}, & \text{if } \lambda \neq 0; \\ \log Y, & \text{if } \lambda = 0. \end{cases} \tag{3}$$

Box-Cox transformation

- As with the other parameters,

Box-Cox transformation

- As with the other parameters, λ can be estimated by maximum likelihood.

Box-Cox transformation

- As with the other parameters, λ can be estimated by maximum likelihood.
- The maximum likelihood estimator of λ does not have a closed-form solution,

Box-Cox transformation

- As with the other parameters, λ can be estimated by maximum likelihood.
- The maximum likelihood estimator of λ does not have a closed-form solution, thus numerical methods are required.

Box-Cox transformation

- As with the other parameters, λ can be estimated by maximum likelihood.
- The maximum likelihood estimator of λ does not have a closed-form solution, thus numerical methods are required.

Box-Cox transformation

Alternative versions for (3) also exist, including:

$$Y^* = \begin{cases} \dfrac{Y^\lambda - 1}{\lambda \dot{Y}^{\lambda-1}}, & \text{if } \lambda \neq 0; \\ \dot{Y} \log Y, & \text{if } \lambda = 0; \end{cases}$$

where $\dot{Y} = \exp\left[\frac{1}{n}\sum_{\ell=1}^{n} \log Y_\ell\right]$ is the geometric mean.

Box-Cox transformation

Alternative versions for (3) also exist, including:

$$Y^* = \begin{cases} \dfrac{Y^\lambda - 1}{\lambda \dot{Y}^{\lambda-1}}, & \text{if } \lambda \neq 0; \\ \dot{Y} \log Y, & \text{if } \lambda = 0; \end{cases}$$

where $\dot{Y} = \exp\left[\frac{1}{n}\sum_{\ell=1}^{n} \log Y_\ell\right]$ is the geometric mean.

Outline

1 Introduction

2 Variance-stabilizing transformations

3 Box-Cox transformation

4 Generalized least squares

5 Application

6 References

Generalized least squares

A linear regression model with non-constant variance can also be fitted using the generalized least squares method.

In this estimation method, the deviations between the observed and expected values of the variable Y_ℓ are multiplied by a weight w_ℓ,

Generalized least squares

A linear regression model with non-constant variance can also be fitted using the generalized least squares method.

In this estimation method, the deviations between the observed and expected values of the variable Y_ℓ are multiplied by a weight w_ℓ, chosen to be inversely proportional to the variance of Y_ℓ, $\ell = 1, 2, \ldots, n$.

Generalized least squares

A linear regression model with non-constant variance can also be fitted using the generalized least squares method.

In this estimation method, the deviations between the observed and expected values of the variable Y_ℓ are multiplied by a weight w_ℓ, chosen to be inversely proportional to the variance of Y_ℓ, $\ell = 1, 2, \ldots, n$.

Generalized least squares

In this method, we assume a linear regression model, with matrix representation given in (2),

Generalized least squares

In this method, we assume a linear regression model, with matrix representation given in (2), satisfying

$$\mathbb{E}(\boldsymbol{\varepsilon}) = \mathbf{0} \text{ and } \text{Var}(\boldsymbol{\varepsilon}) = \sigma^2 \boldsymbol{V}.$$

Generalized least squares

In this method, we assume a linear regression model, with matrix representation given in (2), satisfying

$$\mathbb{E}(\boldsymbol{\varepsilon}) = \mathbf{0} \text{ and } \text{Var}(\boldsymbol{\varepsilon}) = \sigma^2 \boldsymbol{V}.$$

The **generalized least squares estimator** of β is given by:

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top \boldsymbol{V}^{-1} \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{V}^{-1} \boldsymbol{Y}.$$

Generalized least squares

In this method, we assume a linear regression model, with matrix representation given in (2), satisfying

$$\mathbb{E}(\boldsymbol{\varepsilon}) = \mathbf{0} \text{ and } \text{Var}(\boldsymbol{\varepsilon}) = \sigma^2 \boldsymbol{V}.$$

The **generalized least squares estimator** of $\boldsymbol{\beta}$ is given by:

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top \boldsymbol{V}^{-1} \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{V}^{-1} \boldsymbol{Y}.$$

Generalized least squares

When the errors have non-constant variance but are uncorrelated,

$$\sigma^2 \boldsymbol{V} = \begin{bmatrix} 1/w_1 & 0 & \cdots & 0 \\ 0 & 1/w_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1/w_n \end{bmatrix},$$

Generalized least squares

When the errors have non-constant variance but are uncorrelated,

$$\sigma^2 \boldsymbol{V} = \begin{bmatrix} 1/w_1 & 0 & \cdots & 0 \\ 0 & 1/w_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1/w_n \end{bmatrix},$$

the procedure is usually called the weighted least squares method.

Generalized least squares

When the errors have non-constant variance but are uncorrelated,

$$\sigma^2 \boldsymbol{V} = \begin{bmatrix} 1/w_1 & 0 & \cdots & 0 \\ 0 & 1/w_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1/w_n \end{bmatrix},$$

the procedure is usually called the weighted least squares method. And the estimator reduces to: $\widehat{\beta} = (\boldsymbol{X}^\top \boldsymbol{W} \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{W} \boldsymbol{Y}$.

Generalized least squares

When the errors have non-constant variance but are uncorrelated,

$$\sigma^2 \boldsymbol{V} = \begin{bmatrix} 1/w_1 & 0 & \cdots & 0 \\ 0 & 1/w_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1/w_n \end{bmatrix},$$

the procedure is usually called the weighted least squares method. And the estimator reduces to: $\widehat{\beta} = (\boldsymbol{X}^\top \boldsymbol{W} \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{W} \boldsymbol{Y}$.

Generalized least squares

Remarks

The main drawback of this procedure is that the matrix $\boldsymbol{V}$ must be known.

Generalized least squares

Remarks

The main drawback of this procedure is that the matrix $\boldsymbol{V}$ must be known. Otherwise, an initial estimate must be used,

Generalized least squares

Remarks

The main drawback of this procedure is that the matrix $\boldsymbol{V}$ must be known. Otherwise, an initial estimate must be used, followed by iterative re-estimation steps.

Generalized least squares

Remarks

The main drawback of this procedure is that the matrix $\boldsymbol{V}$ must be known. Otherwise, an initial estimate must be used, followed by iterative re-estimation steps.

Outline

1 Introduction

2 Variance-stabilizing transformations

3 Box-Cox transformation

4 Generalized least squares

5 Application

6 References

Application

(Montgomery et al., 2021, p. 180) An electric utility is interested in developing a model relating peak-hour demand (Y, in kW) to total energy usage during the month of August (x_2, in kWh). A scatter plot is given in Figure 1. The least-squares fit is

$$\hat{Y}_\ell = -0.831 + 0.004 x_{\ell 2},$$

$\ell = 1, 2, \ldots, 53.$

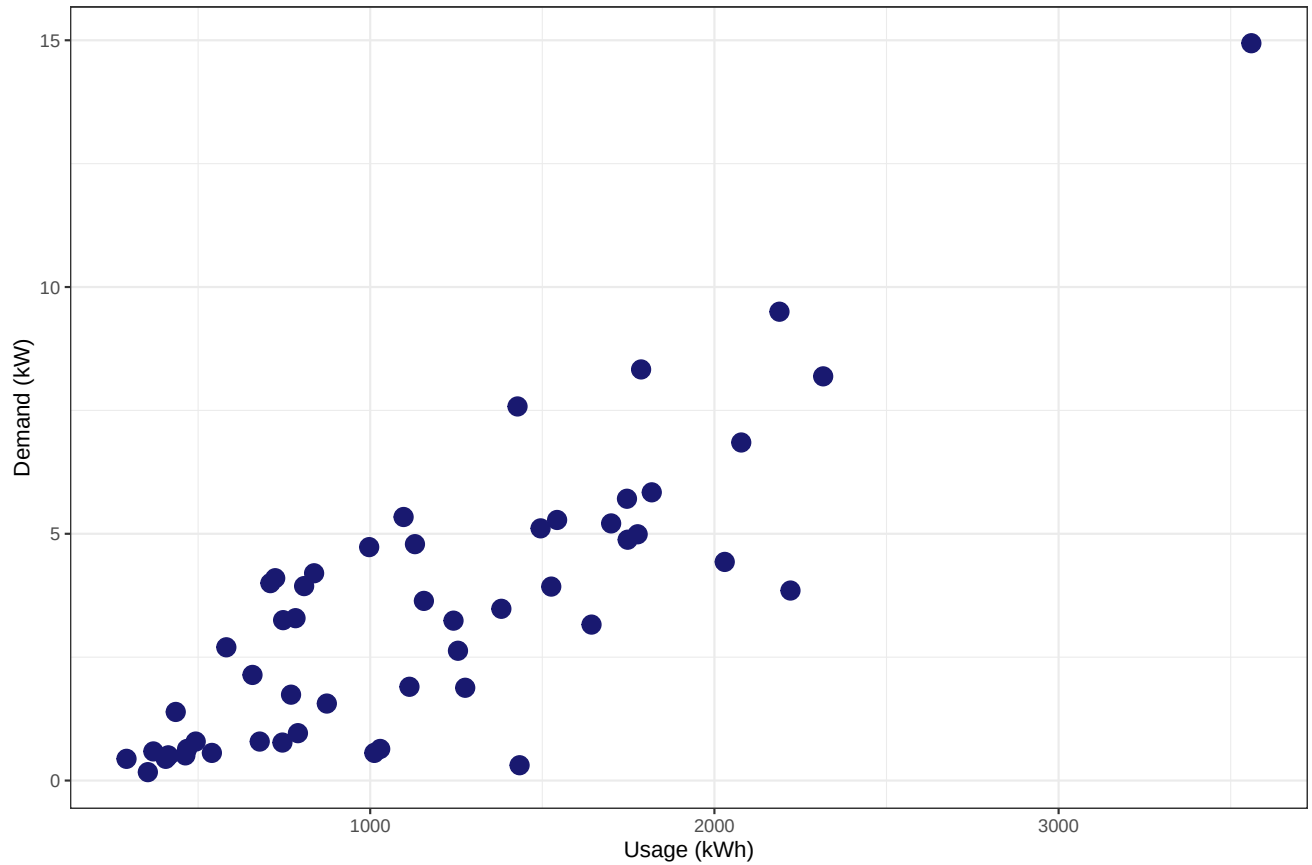


Figure 1: Scatter plot of the energy demand (kW) versus energy usage (kWh).

Example

We also have:

Table 2: Parameter estimates.

Parameter	Estimate	SE	t_0
β_1	-0.831	0.442	-1.882
β_2	0.004	0.001	11.030

Critical region, for $\alpha = 5\%$: $|t_0| > 2.008$, with MSE $= 2.488$.

Example

We also have:

Table 2: Parameter estimates.

Parameter	Estimate	SE	t_0
β_1	-0.831	0.442	-1.882
β_2	0.004	0.001	11.030

Critical region, for $\alpha = 5\%$: $|t_0| > 2.008$, with MSE $= 2.488$.

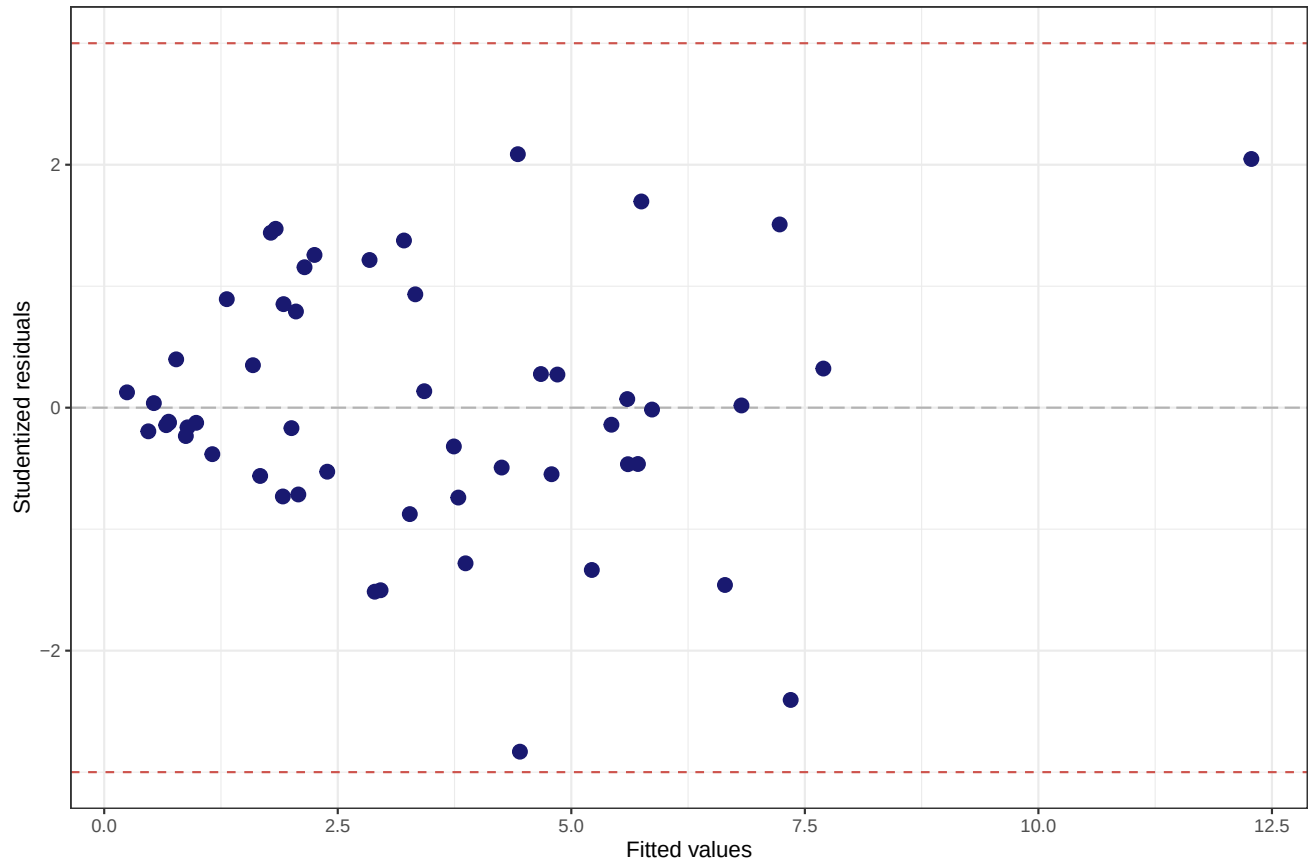


Figure 2: Studentized residuals versus fitted values.

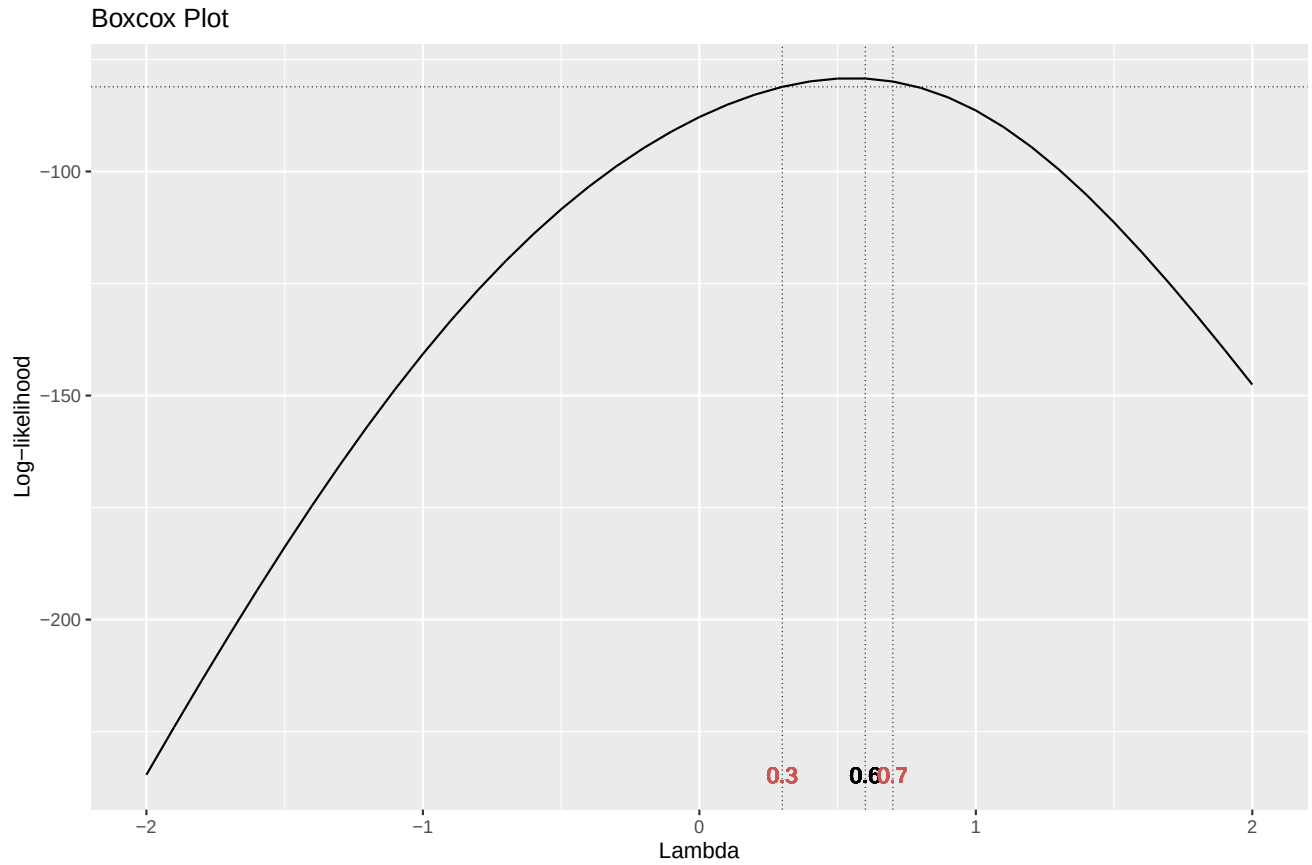


Figure 3: Box-Cox plot, $\lambda = 0.5454$.

Application

The Box-Cox method suggested using (3) with $\lambda = 0.5454$. However, for pedagogical purposes, we will use the following transformation: $Y_\ell^* = \sqrt{Y_\ell}$.

Application

The Box-Cox method suggested using (3) with $\lambda = 0.5454$. However, for pedagogical purposes, we will use the following transformation: $Y_\ell^* = \sqrt{Y_\ell}$.

After fitting the model to the transformed data, we obtain the following estimated model,

$$\hat{Y}_\ell^* = 0.582 + 0.001 x_{\ell 2},$$

$\ell = 1, 2, \ldots, 53$.

Application

The Box-Cox method suggested using (3) with $\lambda = 0.5454$. However, for pedagogical purposes, we will use the following transformation: $Y_\ell^* = \sqrt{Y_\ell}$. After fitting the model to the transformed data, we obtain the following estimated model,

$$\hat{Y}_\ell^* = 0.582 + 0.001 x_{\ell 2},$$

$\ell = 1, 2, \ldots, 53$.

Example

We also have:

Table 3: Parameter estimates.

Parameter	Estimate	SE	t_0
β_1	0.582	0.130	4.481
β_2	0.001	0.001	9.699

Critical region, for $\alpha = 5\%$: $|t_0| > 2.008$, with MSE $= 0.215$.

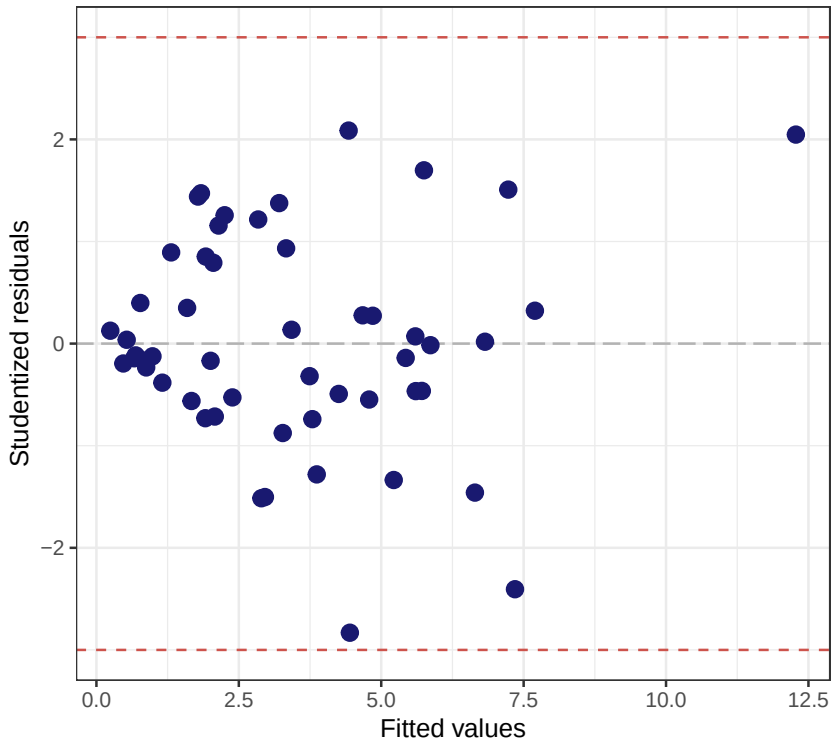
Example

We also have:

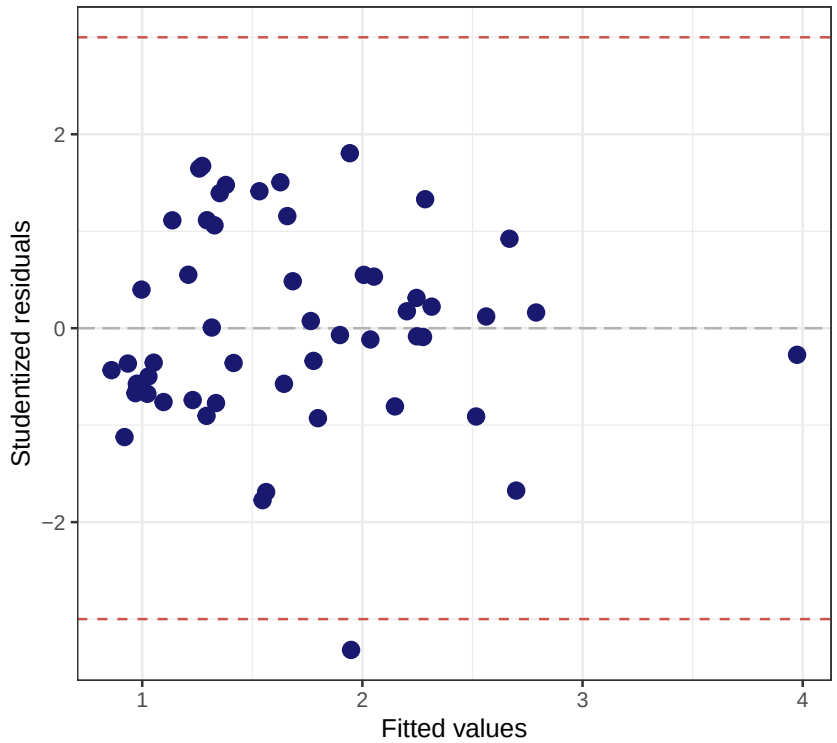
Table 3: Parameter estimates.

Parameter	Estimate	SE	t_0
β_1	0.582	0.130	4.481
β_2	0.001	0.001	9.699

Critical region, for $\alpha = 5\%$: $|t_0| > 2.008$, with MSE $= 0.215$.



(a) Original data.



(b) Transformed data.

Figure 4: Studentized residuals versus fitted values.

Outline

1 Introduction

2 Variance-stabilizing transformations

3 Box-Cox transformation

4 Generalized least squares

5 Application

6 References

References

Box, G. E. P. and Cox, D. R. (1964). An analysis of transformations (with discussion). *Journal of the Royal Statistical Society. Series B (Methodological)*, 26(2):211–252.

Montgomery, D. C., Peck, E. A., and Vining, G. G. (2021). *Introduction to linear regression analysis*. Wiley, New York, 6th edition.

Thank you!

✉ tiago.magalhaes@ufjf.br

ufjf.br/tiago_magalhaes

Department of Statistics, Room 319