STA3010 Regression Analysis

Feng YIN

The Chinese University of Hong Kong (Shenzhen)

yinfeng@cuhk.edu.cn

March 16, 2020

Overview

Variance-Stabilizing

Thus far, we have assumed implicitly that:

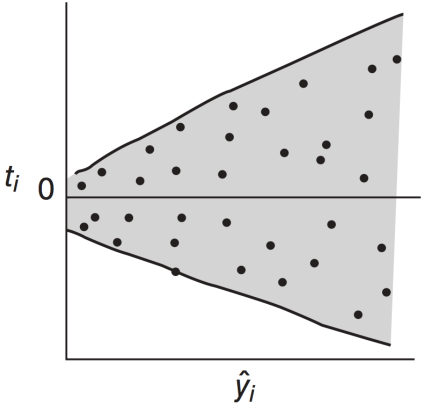
1. The model errors have zero mean and constant variance and they are uncorrelated.
2. The model errors have a normal distribution; This assumption is made in order to conduct hypothesis tests and construct CI. Under this assumption, the errors are independent.

Plots of residuals can be used for detecting violations of these basic regression assumptions. But how to remedy?

Empirical Variance-Stabilizing Transformations

A common reason for violating the “constant variance” assumption lies in that the variance of the output y is functionally related to its mean, $E(y)$.

Empirical Variance-Stabilizing Transformations



Empirical Variance-Stabilizing Transformations

State-of-the-art variance-stabilizing transformations:

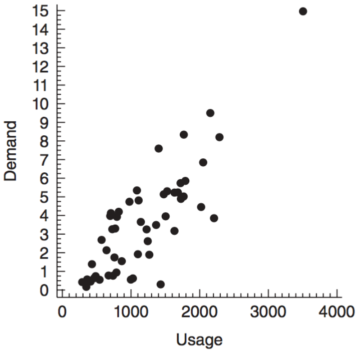
Relationship of σ^2 to $E(y)$	Transformation
$\sigma^2 \propto$ constant	$y' = y$ (no transformation)
$\sigma^2 \propto E(y)$	$y' = \sqrt{y}$ (square root; Poisson data)
$\sigma^2 \propto E(y)[1 - E(y)]$	$y' = \sin^{-1}(\sqrt{y})$ (arcsin; binomial proportions $0 \le y_i \le 1$)
$\sigma^2 \propto [E(y)]^2$	$y' = \ln(y)$(log)
$\sigma^2 \propto [E(y)]^3$	$y' = y^{-1/2}$ (reciprocal square root)
$\sigma^2 \propto [E(y)]^4$	$y' = y^{-1}$(reciprocal)

In practice, a transformation may be selected empirically according to the residual plots.

Empirical Variance-Stabilizing Transformations

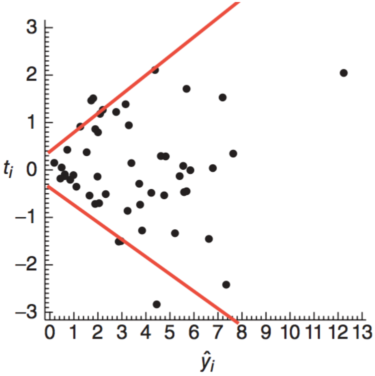
Textbook Example: An electric utility is interested in developing a model to relate the peak-hour demand, y, to the total energy usage during the month, x.

Data set for 53 residential customers collected for one month is shown below:



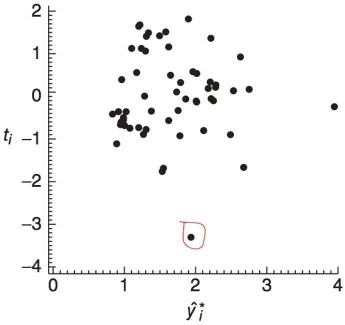
Empirical Variance-Stabilizing Transformations

After applying the simple linear least-squares (LS) fit, we plot the the R-student residuals versus the fitted values $\hat{y}_i$, $i = 1, 2, 3..., n$.



Empirical Variance-Stabilizing Transformations

- Apply $y^* = \sqrt{y}$ as a variance stabilizing transformation and re-do LS fitting.
- The new R-student residuals are plotted against the fitted values.



Analytical Box-Cox Transformation

An important class of transformations for stabilizing the variance is the power transformation/Box-Cox transformation y^λ, where λ is a parameter to be determined.

Due to the difficulties of using y^λ directly, it is modified to be:

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

where $\dot{y}$ is defined to be the geometric mean of the outputs, i.e., $\dot{y} = \left(\prod_{i=1}^n y_i\right)^{1/n}$, which is a scale factor that make the residual sum of squares for models with different values of λ comparable.

After the Box-Cox transformation, we have the new regression model:

$$\mathbf{y}^{(\lambda)} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}'. \tag{2}$$

Analytical Box-Cox Transformation

Complete procedure:

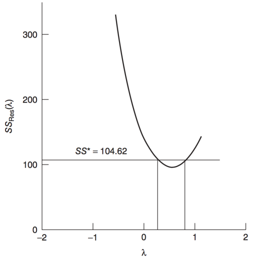
1. Select a few (discrete) λ values, for instance 0:0.5:5
2. For each λ value, do:
    1. Transform the original outputs $\mathbf{y}$ according to Box-Cox transformation, and get transformed output $\mathbf{y}^{(\lambda)}$
    2. Fit the regression model $\mathbf{y}^{(\lambda)} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}^{'}$
    3. Compute the residual sum of squares $SS_{Res}(\lambda)$.
3. Plot $SS_{Res}(\lambda)$ versus λ and find out the λ value that yields the smallest SS_{Res} value.
4. Optional: In the vicinity of the found λ value, repeat the above procedure for refined result.

Analytical Box-Cox Transformation

Textbook Example: An electric utility is interested in developing a model to relate the peak-hour demand, y, to the total energy usage during the month, x.

The values of $SS_{Res}(\lambda)$ versus λ are shown below:

λ	$SS_{\text{Res}}(\lambda)$
–2	34,101.0381
–1	986.0423
–0.5	291.5834
0	134.0940
0.125	118.1982
0.25	107.2057
0.375	100.2561
0.5	96.9495
0.625	97.2889
0.75	101.6869
1	126.8660
2	1,275.5555



Square-root transformation $(\lambda = 0.5)$ is very close to the optimum choice.

Transformations to Linearize A Model

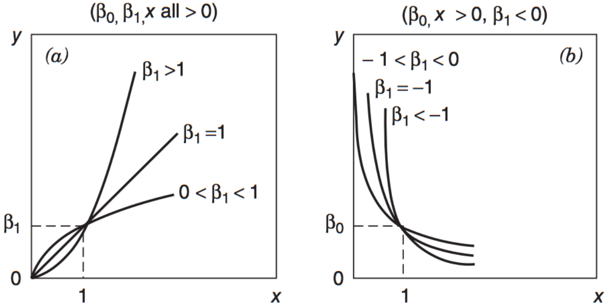
Sometimes, a nonlinear regression model can be linearized through appropriate transformation. Such nonlinear models are called intrinsically or transformably linear.

Some examples:

TABLE 5.4 Linearizable Functions and Corresponding Linear Form

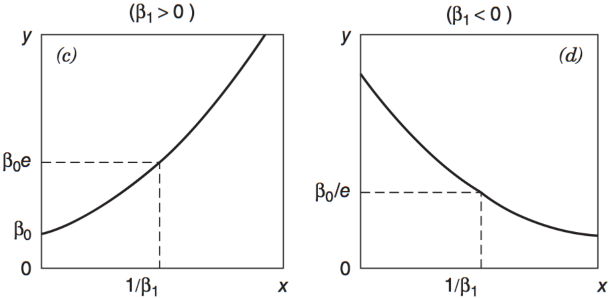
Figure	Linearizable Function	Transformation	Linear Form
5.4*a*, *b*	$y = \beta_0 x^{\beta_1}$	$y' = \log y,\ x' = \log x$	$y' = \log \beta_0 + \beta_1 x'$
5.4*c*, *d*	$y = \beta_0 e^{\beta_1 x}$	$y' = \ln y$	$y' = \ln \beta_0 + \beta_1 x$
5.4g, *h*	$y = \dfrac{x}{\beta_0 x - \beta_1}$	$y' = \dfrac{1}{y},\ x' = \dfrac{1}{x}$	$y' = \beta_0 - \beta_1 x'$

Transformations to Linearize A Model



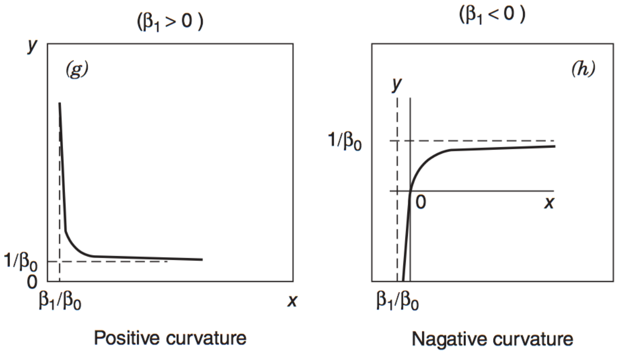
$$y = \beta_0 x^{\beta_1}$$

Transformations to Linearize A Model



$$y = \beta_0 e^{\beta_1 x}$$

Transformations to Linearize A Model



$$y = \frac{x}{\beta_0 x - \beta_1}$$

Transformations to Linearize A Model

One famous transformation:

$$y = \beta_0 e^{\beta_1 x} \varepsilon, \tag{3}$$

where we have multiplicative random error instead of additive random error.

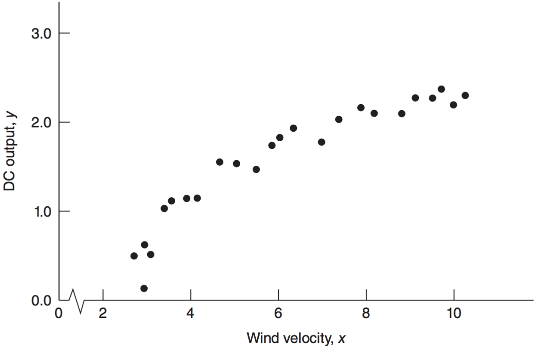
This function is intrinsically linear since it can be transformed to a straight line by logarithmic transformation as

$$\ln y = \ln \beta_0 + \beta_1 x + \ln \varepsilon. \tag{4}$$

If the transformed error terms $\ln \varepsilon$ are normally and independently distributed with mean zero and variance σ^2, then the multiplicative error terms ε in the original model are i.i.d. log-normal distributed.

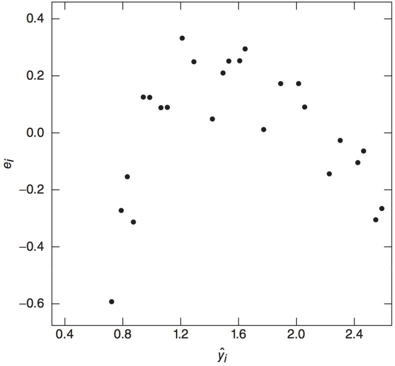
Transformations to Linearize A Model

Textbook example: A research engineer is investigating the use of a windmill to generate electricity. He has collected data on the DC output from his windmill and the corresponding wind velocity.



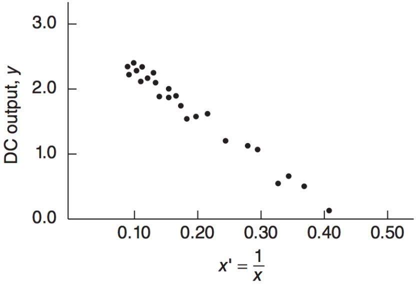
Transformations to Linearize A Model

1. Fitting a straight-line model to the data, yields $\hat{y} = 0.13 + 0.24x$, moreover $R^2 = 0.8745$ and $F_0 = 160.26$.
2. A plot of the residuals versus $\hat{y}_i$ is shown below.



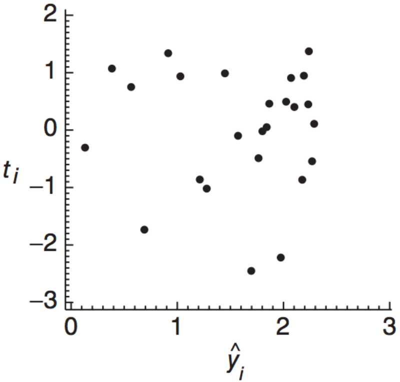
Transformations to Linearize A Model

1. A more reasonable model would be $y = \beta_0 + \beta_1(\frac{1}{x}) + \varepsilon$.
2. Refit the model using LS yields $\hat{y} = 2.98 - 6.93x^{'}$.
3. The summary statistics for the new model are $R^2 = 0.98$, $MS_{Res} = 0.0089$, and $F_0 = 1128.43$.



Transformations to Linearize A Model

A plot of the R-student residuals versus the fitted values, $\hat{y}_i$, is



Summary

1. Variance-Stabilizing
2. Empirical transformations
3. Box-Cox transformation
4. Intrincically linear model
5. Model linearization