Chapter 13 Transformation of the response variable

- Reasons for transformation of the response variable

— To simplify model.
Suppose that

Model 1: $\sqrt[3]{Y} = \beta_0 + \beta_1 X + \varepsilon$

is better than or (almost) as good as

Model 2: $Y = \beta_0 + \beta_1 X \ + \ \beta_{11} X^2 + \ \beta_{111} X^3 + \varepsilon.$

We prefer Model 1 since it is simpler.

— To take error structure into account.
Suppose that a model has multiplicative errors rather than additive errors. For example,

$Y = \alpha X_1^{\beta_1} X_2^{\beta_2} e^{\varepsilon}$, where e = 2.71828 — the natural logarithm base.

Then, transforming the Y into $\ln(Y)$ results in a linear regression model:

$$\ln(Y) = \ln(\alpha) + \beta_1 \ln(X_1) + \beta_1\, ln(X_2) + \varepsilon$$

— To stabilize variance.
Transform Y such that the transformed Y has constant variance.

Q: How? Guess?
A: An important family of transformations is the *family of power transformations*.

- The family of power transformations

Tukey's method (1957):

$$U = \begin{cases} Y^{\lambda} & for\ \lambda \neq 0, \\ \ln(Y) & for\ \lambda = 0. \end{cases}$$

A snag: $\lim_{\lambda \to 0} Y^{\lambda} \ = Y^0 = 1 \neq \ln(Y)$. Thus, $U(\lambda)$ is not continuous at $\lambda = 0$.

Box-Cox method (1964):

$$V = \begin{cases} (Y^{\lambda} - 1)/\lambda & for\ \lambda \neq 0, \\ \ln(Y) & for\ \lambda = 0. \end{cases}$$

Note: $\lim_{\lambda \to 0} \frac{Y^\lambda - 1}{\lambda} = \lim_{\lambda \to 0} \frac{Y^\lambda - Y^0}{\lambda - 0} = \frac{d}{d\lambda}\left(Y^\lambda\right)|_{\lambda=0} = Y^\lambda \ln(Y)|_{\lambda=0} = \ln(Y)$. Thus $V(\lambda)$ is a continuous function of λ.

Let $RSS(\lambda)$ is the residual sum of squares when V is the response. Then under the assumptions for the model, the log-likelihood function

$$L(\lambda) = -\frac{n}{2}\ln\left(\frac{RSS(\lambda)}{n}\right) + (\lambda - 1)\sum_{i=1}^{n} ln(Y_i).$$

To find λ, plot $L(\lambda)$ vs. λ and choose the value of λ that maximizes $L(\lambda)$, denoted by $\hat{\lambda}$.

- Some notes

(1) Transforming the response can make the model harder to interpret. So, we do not want to do it unless it is really necessary, which can be checked by a $100(1-\alpha)\%$ CI for λ:

$$\{\lambda: L(\lambda) > L(\hat{\lambda}) - \tfrac{1}{2}\chi_1^2(1-\alpha)\}$$

— A transformation of the response is really necessary if 1 is outside the CI.

(2) For convenience, we can use V to determine λ and carry out analysis on U instead of V since $\frac{Y^\lambda - 1}{\lambda} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \varepsilon \Leftrightarrow Y^\lambda = (1 + \lambda\beta_0) + \lambda\beta_1 X_1 + \cdots + \lambda\beta_k X + \lambda\varepsilon$.

(3) If explaining the model is important, you should round $\hat{\lambda}$ to the nearest interpretable value, for example, round $\hat{\lambda} = 0.42$ to ½.

(4) To use a power transformation, all $Y_i^{'}s$ must be positive. If some $Y_i^{'}s$ are negative, we can work with $Y_i + c$, where c is chosen such that $Y_i + c > 0$ for all i.

(5) For observations on a positive Y, $Y_1, \cdots, Y_n$, if $\frac{\max(Y_1, \cdots, Y)}{min(Y_1, \cdots, Y_n)} \geq 10$, a transformation on Y is likely to be effective.

(6) There is no guarantee that even the best choice of λ will produce a transformed set of values that adequately conform to the assumptions.

- Important plots to check the assumptions for errors

Since transformations on the response variable change error structure, it is very important to check the assumptions for errors in the transformed response: $\boldsymbol{\varepsilon} \sim N_n(0, \sigma^2 \boldsymbol{I}_n)$.

(a) The Q-Q plot of residuals to check normality.
(b) Plot residual versus fitted value.
(c) Plot residual against each predictor.
(d) Plot X_i against X_j for $i \neq j$. If two predictors X_i and X_j are highly correlated, generally it is unnecessary to include both variables in the model.

Ex. 13.1 In an experiment to study the effect of filler and naphthenic oil levels on mooney viscosity MS_4 at 100°C, the following data are observed. Find the best Box-Cox transformation on *Y* for the data and assess its effectiveness.

Naphthenic Oil, phr, *p*	Filler, phr, *f*	Mooney Viscosity, *Y*
0	0	26
0	12	38
0	24	50
0	36	76
0	48	108
0	60	157
10	0	17
10	12	26
10	24	37
10	36	53
10	48	83
10	60	124
20	0	13
20	12	20
20	24	27
20	36	37
20	48	57
20	60	87
30	12	15
30	24	22
30	36	27
30	48	41
30	60	63