Lecture Summary: Feb. 10, 2023

• Remedial measures

1. Transformation of x: For nonlinear association.

2. Transformation of Y: For non-normality/unequal variance.

Box-Cox transformation

This is a collection of transformations depending on a "tuning parameter", λ. The definition is the following:

$$Y_i' = \begin{cases} K_1(Y_i^\lambda - 1), & \lambda \neq 0, \\ K_2 \log(Y_i), & \lambda = 0, \end{cases}$$

where K_1, K_2 are two numbers computed from the data, with K_1 depending on λ:

$$K_2 = (Y_1 Y_2 \dots Y_n)^{1/n} = e^{\overline{\log Y}}, \quad K_1 = \frac{1}{\lambda K_2^{\lambda-1}},$$

where $\overline{\log Y} = n^{-1} \sum_{i=1}^n \log(Y_i)$ (log means natural logarithm).

How does it work?

(i) For each λ, find the SSE of regression of Y' on x (i.e., Y' is the response variable, x is the predictor variable).

(ii) The SSE depends on λ. Find the λ whose SSE is the smallest, which corresponds to the optimal transformation.