Methods for Choosing Transformations

Box-Cox Transformation of Y

The Box-Cox family of transformations of Y is given by:

$$Y_i^{(\lambda)} = \alpha + \beta_1 X_{i1} + \cdots + \beta_k X_{ik} + \varepsilon_i,$$

for $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$, and

$$Y_i^{(\lambda)} = \left\{ \begin{array}{cc} (Y_i^\lambda - 1)/\lambda & \text{for } \lambda \neq 0 \\ \log Y_i & \text{for } \lambda = 0 \end{array} \right\}$$

where $Y_i > 0$. For a given value of λ, the conditional maximized log-likelihood is

$$\log(\alpha, \beta_1, ..., \beta_k, \sigma_\varepsilon^2|\lambda) = -\frac{n}{2}(1 + \log 2\pi) - \frac{n}{2}\log \widehat{\sigma}_\varepsilon^2(\lambda) + (\lambda - 1)\sum_{i=1}^{n} \log Y_i,$$

where $\widehat{\sigma}_\varepsilon^2(\lambda) = \sum E_i^2(\lambda)/n$ and the $E_i(\lambda)$ are the residuals from the regression of $Y^{(\lambda)}$ on the Xs. Often the profile log-likelihood (maximized $\log L$) is plotted as a function of λ to locate the maximum value, or a well-known value close to the MLE of λ. We can test for the need to transform Y ($H_0 : \lambda = 1$) with the likelihood-ratio statistic:

$$G_0^2 = -2[\log L(\lambda = 1) - \log L(\lambda = \widehat{\lambda})],$$

which has an asymptotic χ^2 distribution with 1 degree of freedom under H_0. As shown in the text, a confidence interval for λ can be obtained by inverting the likelihood-ratio statistic and finding λ values that satisfy

$$\log L(\lambda) > \log L(\lambda = \widehat{\lambda}) - 1.92.$$

An approximate score test was proposed by Atkinson (1985) that uses a constructed variable, thus creating a test for the need for transformation and also a series of useful plots.