Theory of Linear Models
Exercises 4 Solutions

10 February 2020

1. In the variance components model given in lectures, write down the set of values which can be take by the variance component σ_1^2.

   From the form of the matrix $\mathbf{V}$ in lectures, it is clear that $\sigma_1^2 + \sigma^2 \geq 0$ and so $\sigma_1^2 \geq -\sigma^2$, so negative variance components are possible.

2. Write down the form of a generalized linear model with a normal distribution and a log link function. Explain the difference between this and a Box-Cox transformation with $\lambda = 0$.

   The GLM is $Y_i \sim N(\mu_i, \sigma^2)$, where $\log \mu_i = \mathbf{x}_i' \boldsymbol{\beta}$. Here the data have a normal distribution and so, for example, can take negative values, even though they come from a distribution whose mean must be positive. In the Box-Cox transformation, the data must be positive and come from a positively skewed distribution.

3. Consider the two candidate models

$$Y_i = \theta_0 e^{\theta_1 x_i} + \epsilon_i$$

and

$$\log Y_i = \log \theta_0 + \theta_1 x_i + \epsilon_i,$$

where $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$ in each case.

(a) Explain the similarities and differences between these two models.

The models have the same relationship between the mean response and the explanatory variables, but a different assumed distribution, e.g. the first model assumes constant variance, while the second assumes that the variance increases with the mean.

(b) How should an experimenter decide between them?

Using prior knowledge (e.g. that the data must be positive) or residuals from fitting to decide what is the more appropriate variance structure.

(c) Suggest a more general model that could be used instead of either of these.

$$Y_i^{(\lambda)} = \left(\theta_0 e^{\theta_1 x_i}\right)^{(\lambda)} + \epsilon_i.$$