Math 5366 Homework 29

1. Import the file `math5305Lab6Data.txt`, whose columns are the variables Y, X_1, X_2, and X_3. In Homework 27, we saw that the model

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \epsilon_i$$

does not satisfy the assumption $\epsilon_i \sim N(0, \sigma^2)$, $i = 1, \ldots, n$. To remedy this, use SAS to perform a Box-Cox transformation of Y by defining $\tilde{Y}_i = (Y_i^\lambda - 1)/\lambda$, for $i = 1, \ldots, 100$.

   (a) Fit the model

$$\tilde{Y}_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \epsilon_i,$$

   and let $\hat{\tilde{Y}}$ and $\tilde{e}$ be the predicted values and residuals for this transformed model.

   (b) Plot $\tilde{Y}$ vs. $\hat{\tilde{Y}}$ and $\tilde{e}$ vs. $\hat{\tilde{Y}}$. Does curvature appear to exist in the transformed model?

   (c) Investigate normality of the errors for the transformed model.

   (d) Investigate constancy of error variance for the transformed model.

   (e) Do the errors for the transformed model appear to satisfy the assumptions of normality and constant error variance? How do your results compare to those from Homework 23?

2. The file `math5305Lab7Data.txt` contains data for the variables $Y, X_1, X_2, \ldots, X_{40}$. Perform a stepwise regression on this data set using SAS. (Hints: It may be helpful to use the "Import Data" option in the "File" menu to import this data. Also, make sure to specify in your `glmselect` procedure which variables are `class` variables. Finally, it may be convenient to use R to generate the `model` statement for this procedure.)