Stat 6021: Homework Set 5

1. For this question, we will use the `cornnit` data set from the `faraway` package. Be sure to install and load the `faraway` package first, and then load the data set. The data explore the relationship between corn yield (bushels per acre) and nitrogen (pounds per acre) fertilizer application in a study carried out in Wisconsin.

   (a) What is the response variable and predictor for this study? Create a scatterplot of the data, and interpret the scatterplot.

   (b) Fit a linear regression without any transformations. Create the corresponding residual plot. Based only on the residual plot, what transformation will you consider first? Be sure to explain your reason.

   (c) Create a Box Cox plot for the profile loglikelihoods. How does this plot aid in your data transformation?

   (d) Perform the necessary transformation to the data. Re fit the regression with the transformed variable(s) and assess the regression assumptions. You may have to apply transformations a number of times. Be sure to explain the reason behind each of your transformations. Perform the needed transformations until the regression assumptions are met. What is the regression equation that you will use?

   Note: in part 1d, there are a number of solutions that will work. You must clearly document your reasons for each of your transformations.

2. (No R required) A chemist studied the concentration of a solution, y, over time, x, by fitting a simple linear regression. The scatterplot of the dataset, and the residual plot from the regression model are shown in Figure 1.

   (a) Based only on Figure 1, would you recommend transforming the predictor, x, or the response, y, first? Briefly explain your choice.

   (b) The profile log-likelihoods for the parameter, λ, of the Box-Cox power transformation, is shown in Figure 2. Your classmate says that you should apply a log transformation to the response variable first. Do you agree with your classmate? Be sure to justify your answer.

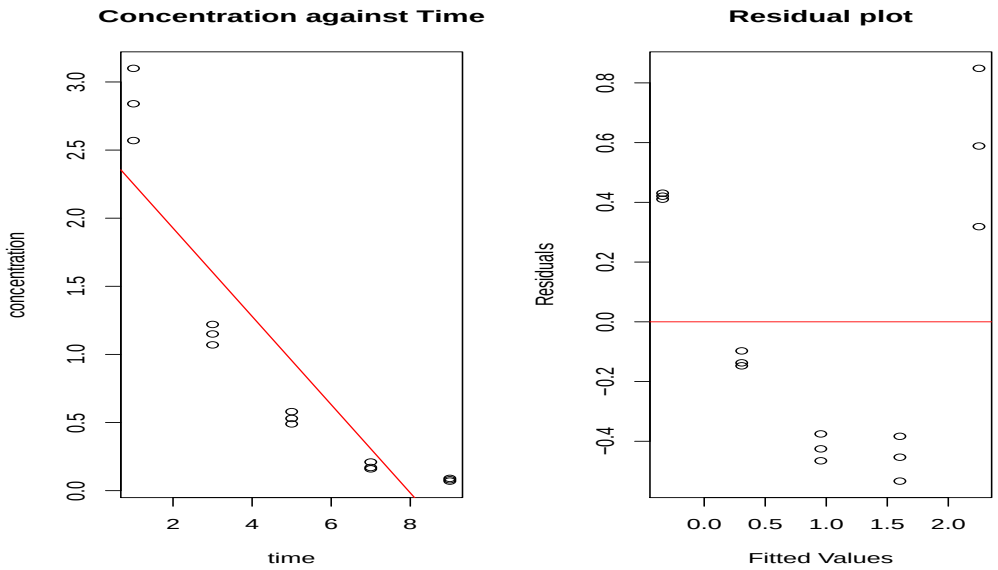

Figure 1: Scatterplot of Concentration of Solution against Time (left). Residual Plot from SLR (right)

(c) Your classmate is adament on applying the log transformation to the response variable, and fits the regression model. The R output is shown in Figure 3. Write down the estimated regression equation for this model. How do we interpret the regression coefficients $\hat{\beta}_1$ and $\hat{\beta}_0$ in context?

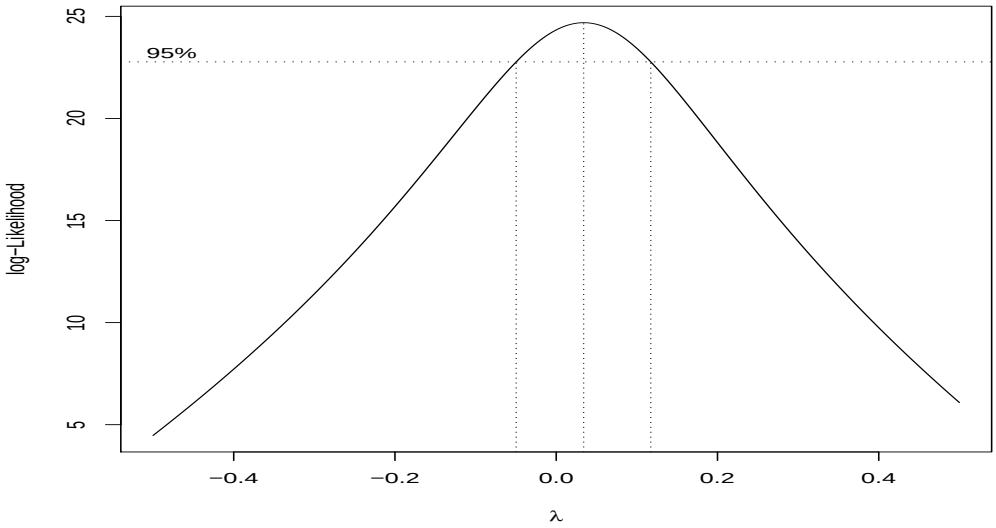

Figure 2: Profile Log-likelihoods for λ.

```
Call:
lm(formula = l.conc ~ time)

Residuals:
     Min       1Q   Median       3Q      Max
-0.19102 -0.10228  0.01569  0.07716  0.19699

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.50792    0.06028   25.01 2.22e-12 ***
time        -0.44993    0.01049  -42.88 2.19e-15 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.115 on 13 degrees of freedom
Multiple R-squared:  0.993,     Adjusted R-squared:  0.9924
F-statistic:  1838 on 1 and 13 DF,  p-value: 2.188e-15
```

Figure 3: R Output after Transforming Response Variable.