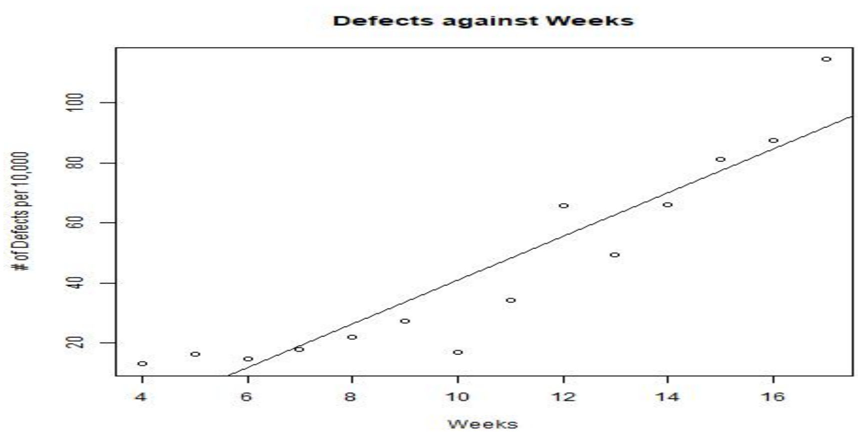
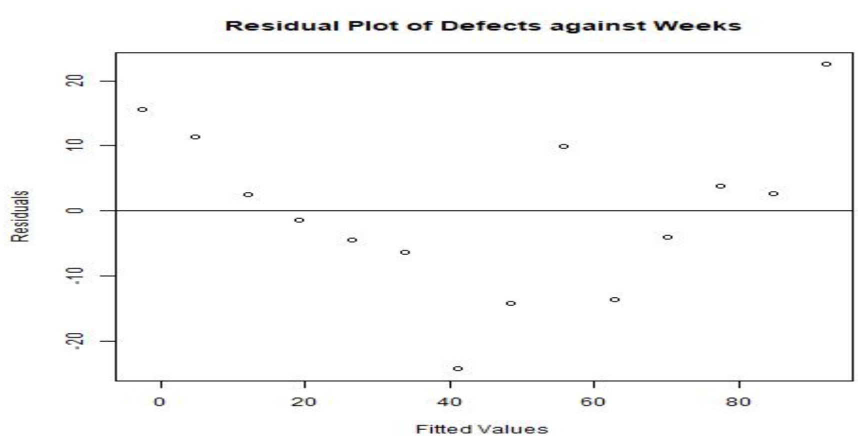
Stat 6021: Guided Question Set 3 Solutions

1. Note that there are portions of the scatterplot where the data points fall on only one side of the regression equation, so the scatterplot has a curved pattern. The linearity assumption is not met (or one can say the residuals are not 0, on average, for each value of the predictor). It is difficult to evaluate the constant variance assumption in this particular scatterplot since we cannot evaluate the spread of the response variable for each value of the predictor.



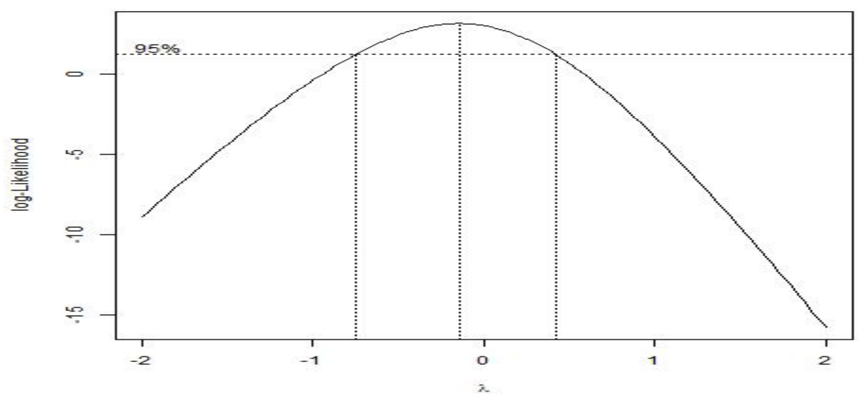
2. The residual plot has a curved pattern, so the linearity assumption is not met. It is difficult to evaluate the constant variance assumption in this particular scatterplot since we cannot evaluate the spread of the residuals for each value of the fitted response.



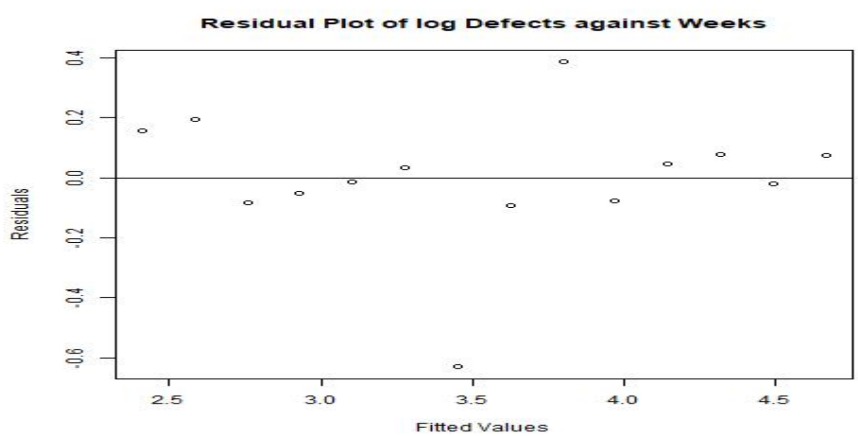
3. Both plots indicate an issue with the linearity of the relationship, so we definitely need to transform at least the predictor. Based on the plots, it is difficult to visualize if the constant variance assumption is met.

4. The Box Cox can be used to help decide how to transform the response variable. We transform the response variable when the constant variance assumption is not met. Given that the residual plot, for this data set, doesn't help us evaluate this assumption, the Box Cox plot will be especially important to check.

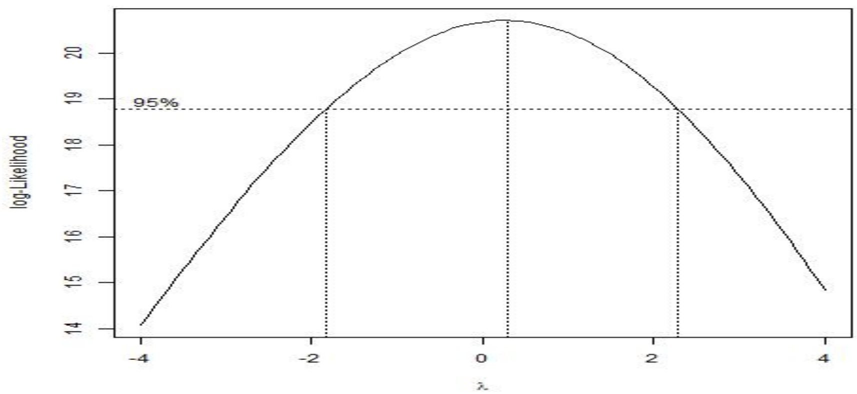
5. Since the 95% confidence interval does not contain $\lambda = 1$, we need to transform the response. A log transformation is preferable since regression coefficients still have interpretive value. Since $\lambda = 0$ lies inside the interval, we can apply a log transformation to the response variable.



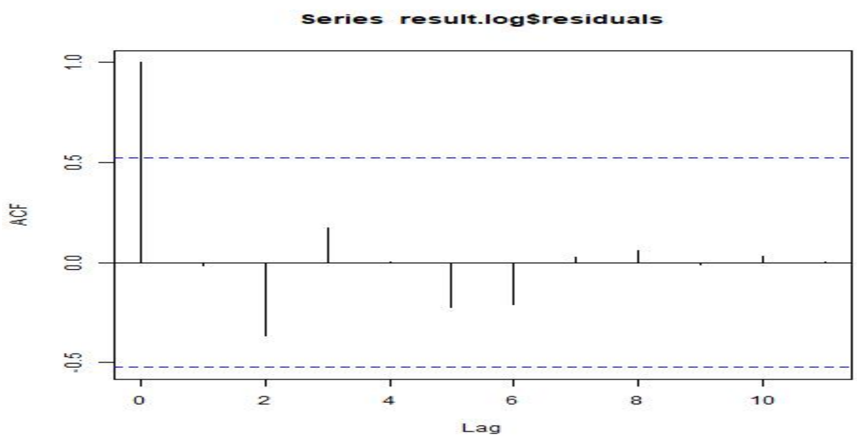
6. Residual plot looks a lot better. The curved pattern is much less pronounced. Given the small number of observations, it is still difficult to evaluate the constant variance assumption based on this plot. Again, this residual plot is not useful in evaluating whether the constant variance assumption is met, so a Box Cox plot is produced.



Notice that $\lambda = 1$ lies inside the interval, so we no longer need to transform the response variable. The log transformation was enough.



7. The autocorrelations from lag 1 onwards are all insignificant, so the independent error assumption is not violated.



8. The pattern we see in the QQ plot indicates the tails are heavier than in a normal distribution.

   Note: So our model is $\log y = \beta_0 + \beta_1 x + \epsilon$.