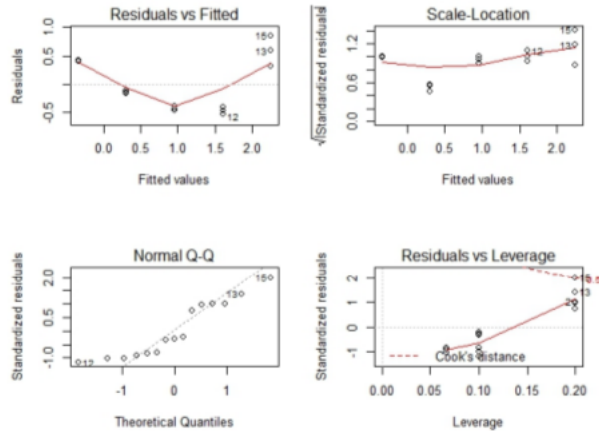
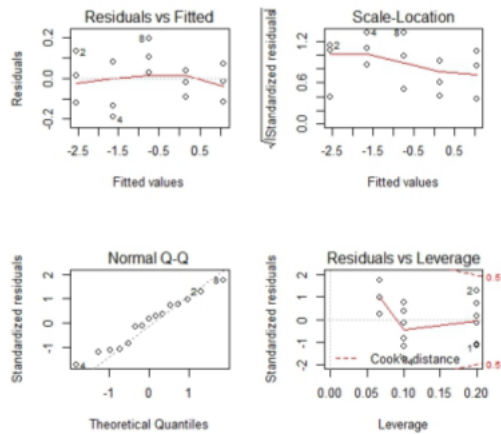


STAT 217: Quiz 22

- A chemist studied the concentration of a solution over time. Fifteen identical solutions were prepared. The 15 solutions were then randomly divided into five sets of three, and the five sets were measured, respectively after 1, 3, 5, 7, and 9 hours. A simple linear regression model was fit to these data and the diagnostic plots are found below.



The chemist noticed a nonlinear relationship between these two variables and decides to rerun his analysis by taking the log of the concentration and using $\log(\text{concentration})$ as the response variable. The diagnostic plots for this model are found below.

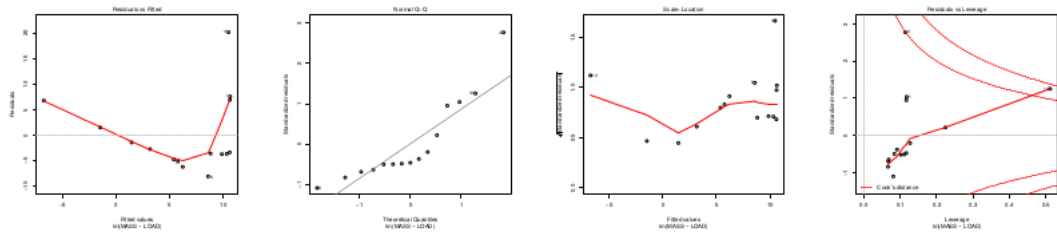


- Why was a log transformation on *concentration* considered for this example?
- Did the log transformation on *concentration* improve the model assumptions? Briefly justify.

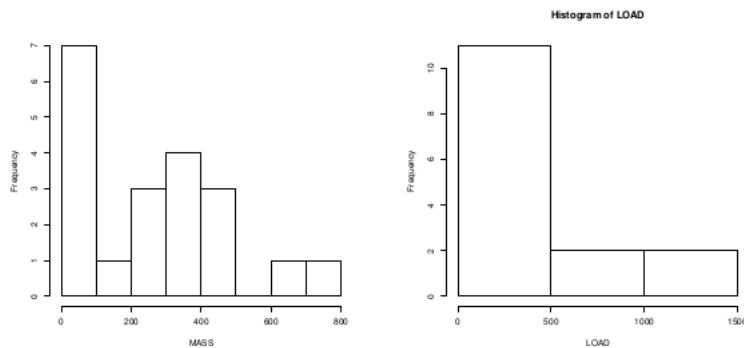
2. Biological Pest Control. In a study of the effectiveness of biological control of the exotic weed tansy ragwort, researchers manipulated the exposure to the ragwort flea beetle on 15 plots that had been planted with high density of ragwort. Harvesting the plots the next season, they measured the average dry mass of ragwort remaining (grams/plant) and the flea beetle load (beetles/gram of ragwort dry mass) to see if the ragwort plants in plots with high flea beetle loads were smaller as a result of herbivory by the beetles.

(a) What is the response variable, and what is the explanatory variable?

- (b) A linear model was fit. Look at the diagnostic plots. Do you think a transformation is needed? Why or why not?



- (c) Based on the two histograms below, what transformation(s) do you suggest? Why?



- (d) Below is the summary table of the linear regression fit on the transformed scale. Which variable(s) was (were) log transformed?

```
Call:
lm(formula = log(MASS) ~ log(LOAD), data = pest[-15, ])
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-2.98400 -0.87993 -0.04895  1.21008  2.00499
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   7.3410     1.3080   5.612 0.000114
log(LOAD)    -1.5108     0.2565  -5.890 7.37e-05
---
```

```
Residual standard error: 1.515 on 12 degrees of freedom
Multiple R-squared:  0.743, Adjusted R-squared:  0.7216
F-statistic: 34.69 on 1 and 12 DF, p-value: 7.369e-05
```

- (e) Write out the estimated regression equation for this model.
- (f) Now write out the equation for the TRUE regression line for this model.
- (g) One of the observed average dry mass of the remaining ragwort was zero. I had to drop this observation from the analysis using log transformations. Why did I have to do so?