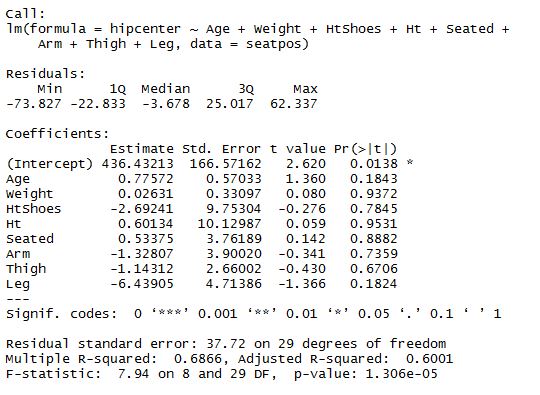
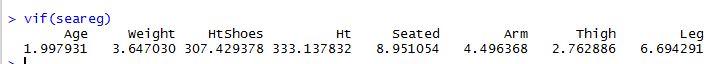
**HW 3**

**Problem 2**

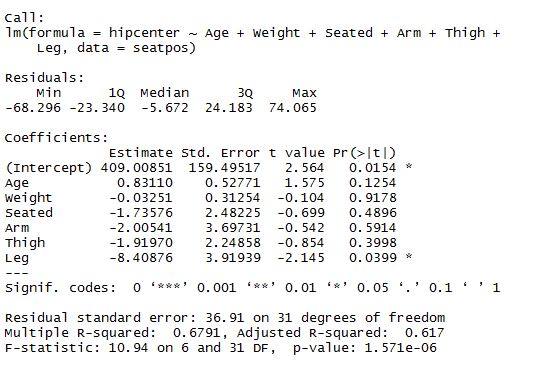
1. Summary of the regression results:



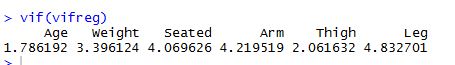
1. No variable appear to be significant based on the individual t-tests for their coefficients. The only thing that’s significant is the intercept. The overall F-test is significant, with p-value <.05.
2. According to the VIFs shown below, HtShoes and Ht have VIFs larger than 10, indicating a possible problem of collinearity.



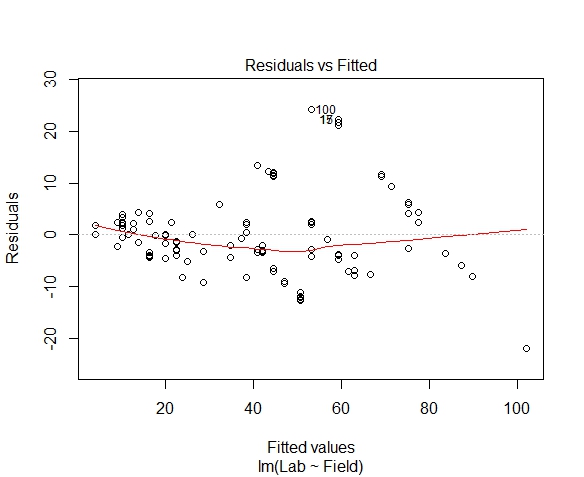
1. Summary of regression results moving variables with VIFs larger than 10

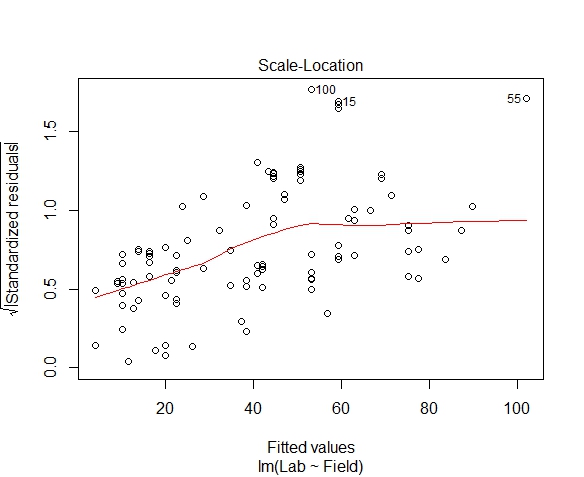


1. For the model in (d), Leg is significant based on individual t-test. The overall F-test is also significant, with p-value <.05.
2. VIFs for the new model: there are not any variables with VIFs larger than 10. Also, the VIFs become smaller compared with the results in (c).

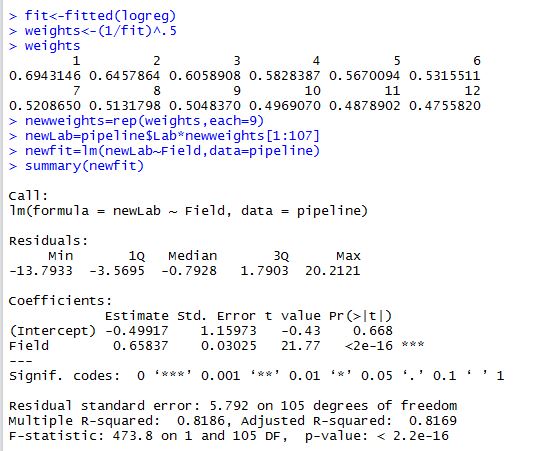


**Problem 3**

1. According to the Residual vs. Fitted and the Scale-Location plots shown below, there is probably a problem with constant variance, in that the points do not equally spread vertically in the Residual vs Fitted plot, and the lines are not flat in both plots. ****



1. The R code, weights, and the regression summary are shown below:



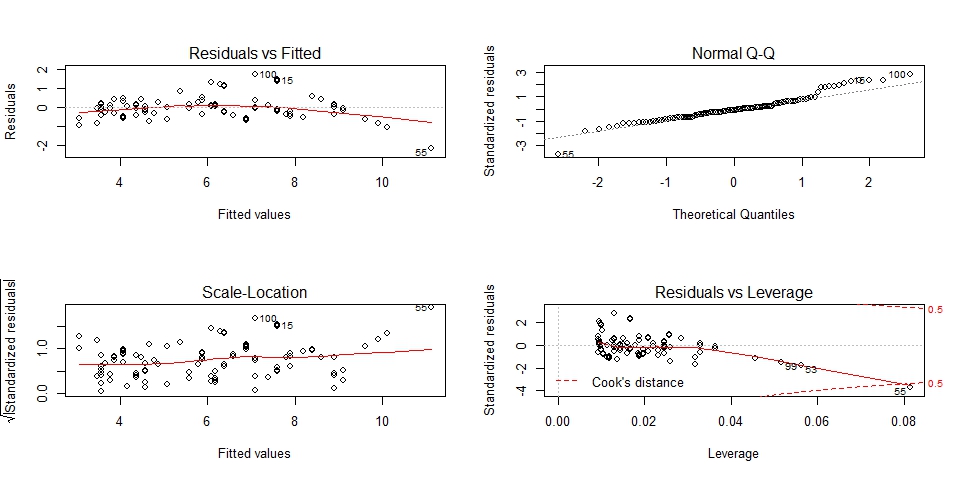
1. I fit the regression model by transforming in different ways the dependent variable, and the diagnostic plots obtained are shown below. Based on the diagnostic plots, I think the model looks best in terms of linearity and constant variance when the square root transformation is conducted on the response variable.

In the Residual vs. Fitted plot under “\*Square root”, which is under “Transformation on response variable”, we can see that the line is roughly flat, and points spread roughly equally vertically. The roughly flat line in Scale-Location is another indication of constant variance.

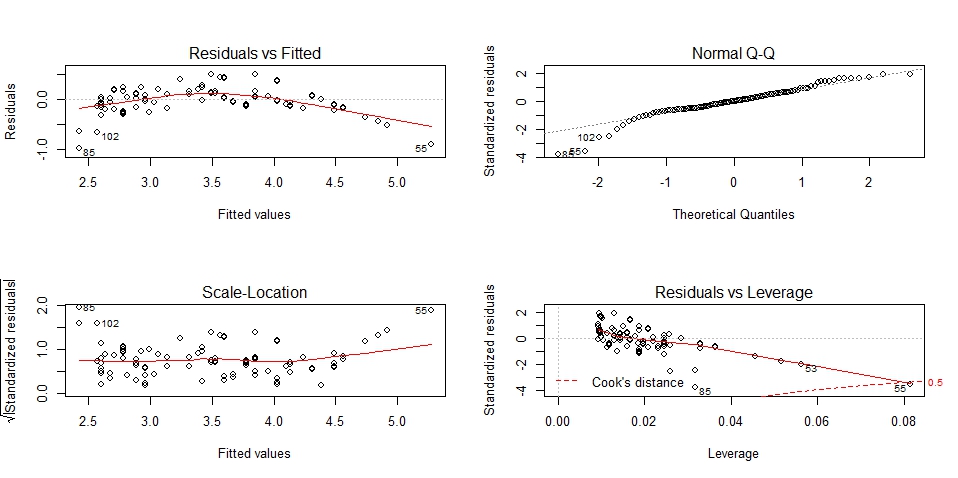
Compared with this model, other models are a lot farther from being satisfactory.

**Transformation on response variable**

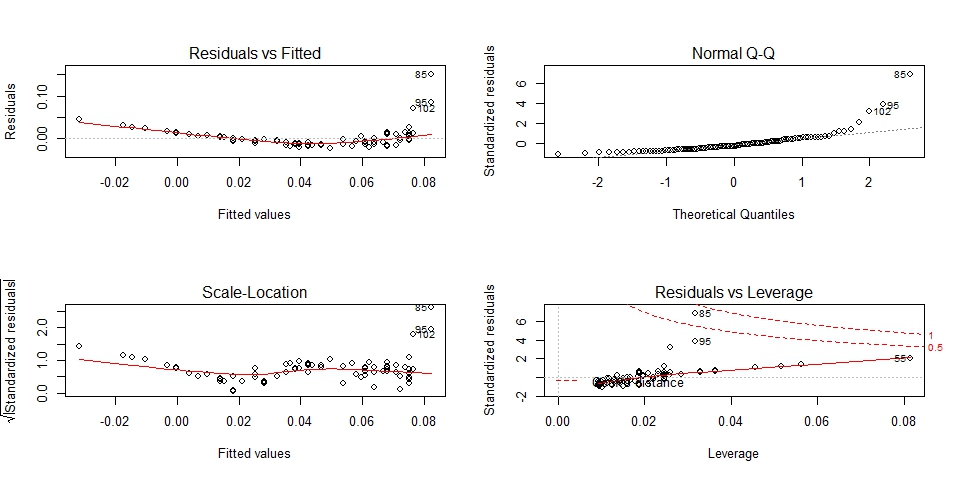
**\*Square root**



**\*Log**

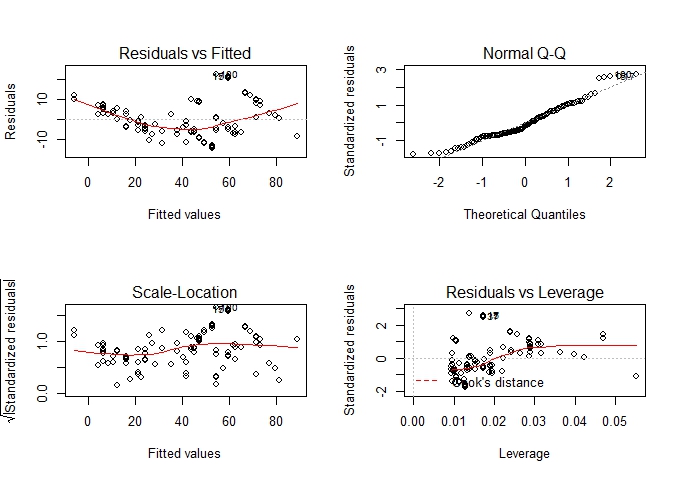


**\*Inverse**

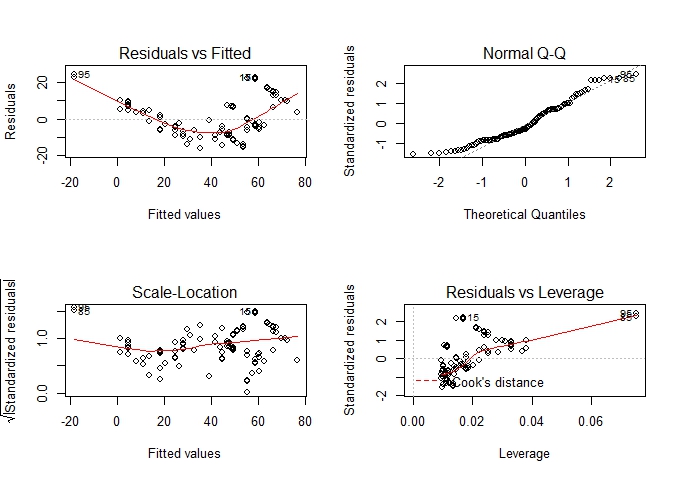


**Transformation on predictor:**

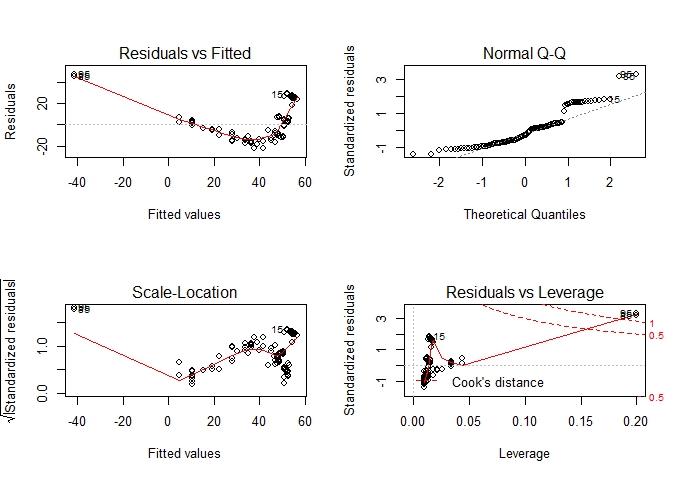
**\*Square root**



**\*Log**



**\*Inverse**



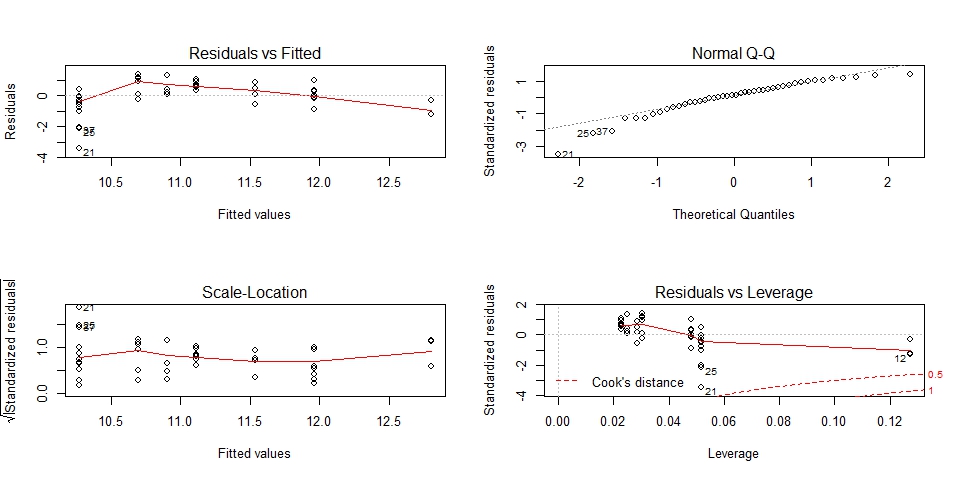
**Problem 4**

1. Diagnostic plots of models after various transformations are shown below. I think the model that looks the best in terms of linearity and constant variance is the model where we take the square root of the response variable and then fit the model.

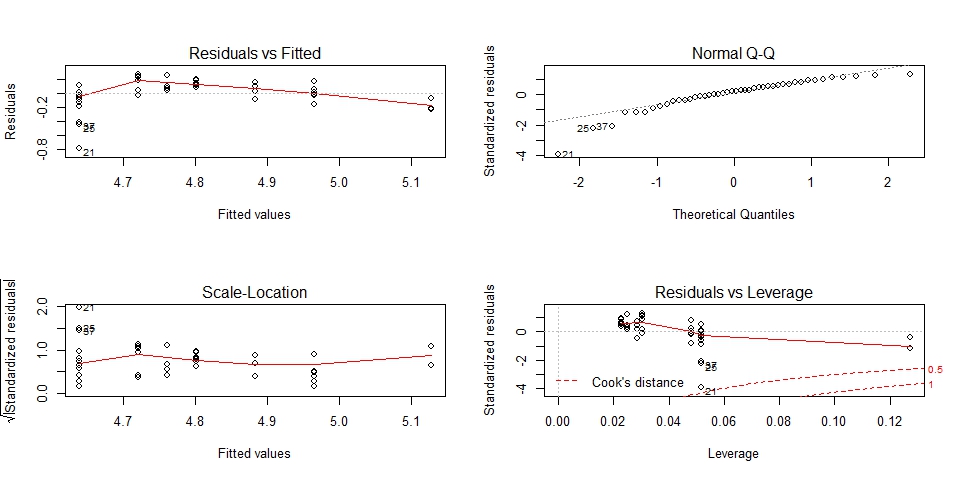
By comparing the 6 plots where different ways of transformations were conducted, the plots of the model where the square root was taken for the response variable has a relatively flat line in the Residual vs Fitted plot, and the points are roughly spreading evenly vertically, indicating linearity and constant variance.

**\*Response variable transformation**

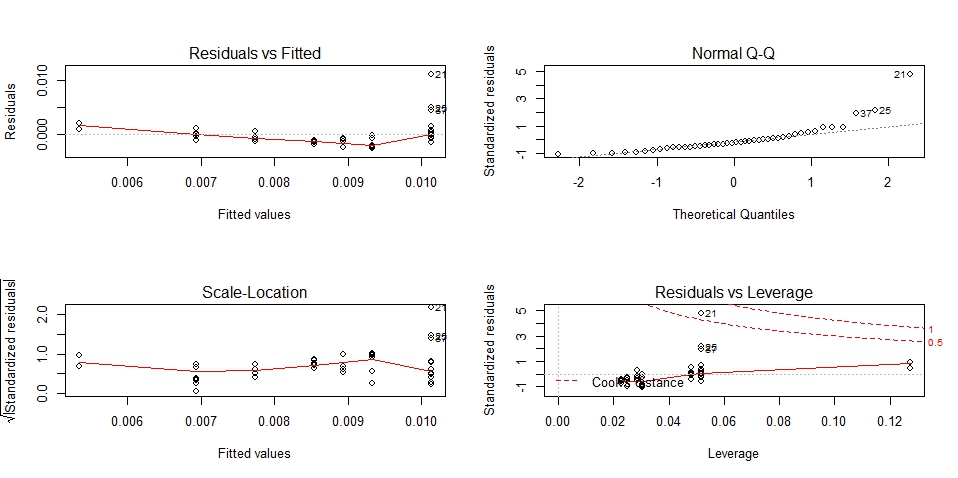
**Square root: (the best one)**



**Log:**

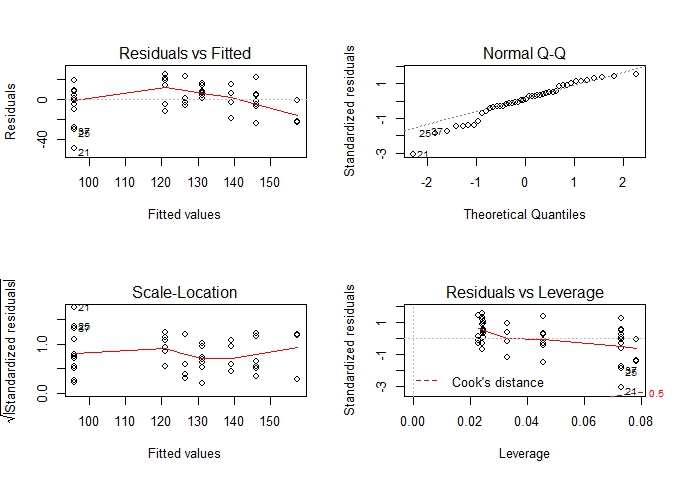


**Inverse:**

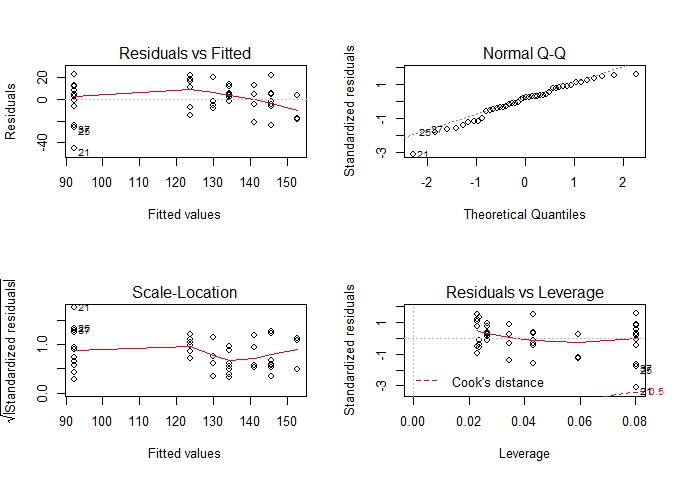


**\*Predictors transformation:**

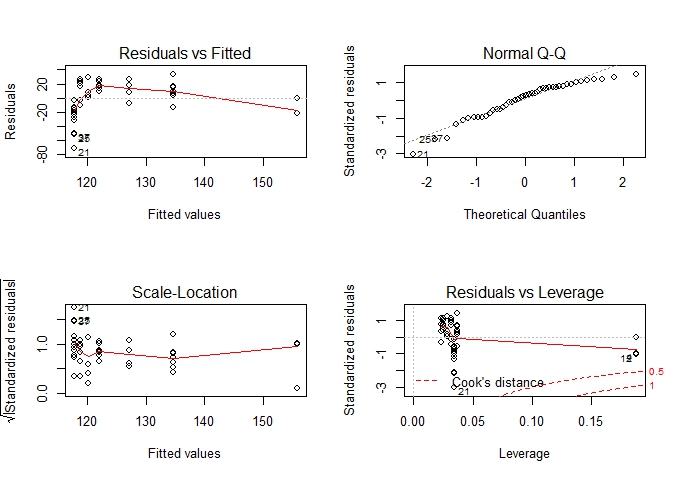
**Square root:**



**Log (10+…)**



**Nitrogen^2**



1. To test lack-of-fit for the best model picked from (a), I fit the model again using the square root of “corn” as response variable, and set the predictor “nitrogen” as factor, followed by an F-test between the best model in (a) and this new model. The F-test results are shown below, with a p-value smaller than 0.05, so the new model with bigger df but smaller RSS is actually better than the “best model” obtained from (a).

