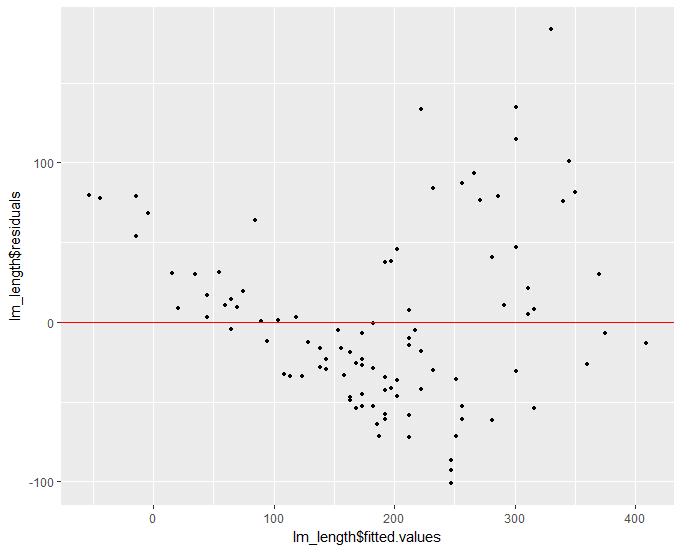
Lab 8

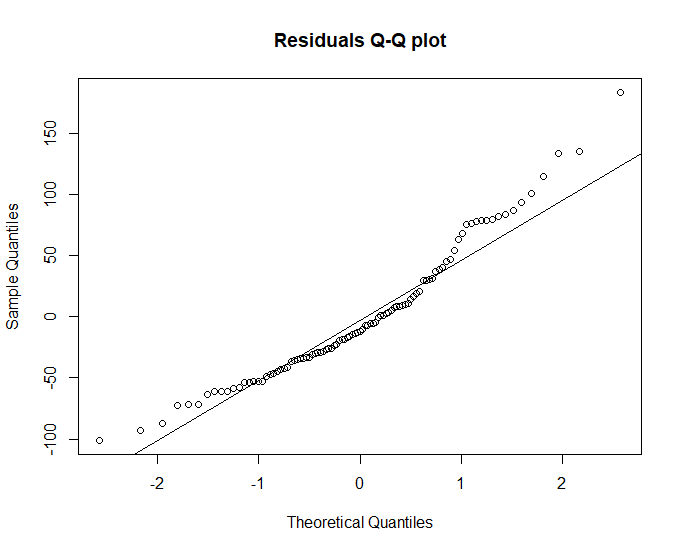
Kyle Salitrik

997543474

# Part 1

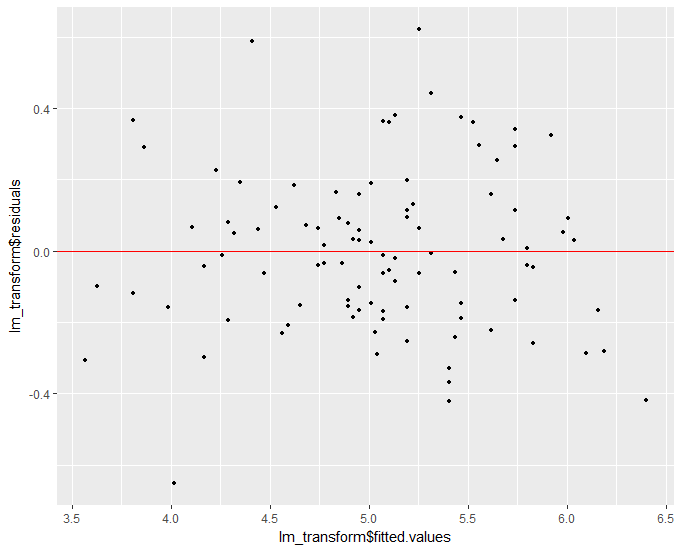
After plotting the residuals vs. the fitted values (shown below) it is clear that there exists a trend of at least a parabolic polynomial within the data. It also looks like the variance is non-constant but that cannot be purely determined from here. To improve the fit of the model, adding in higher order polynomial terms may help.

Taking a look at the Q-Q plot of the residuals shown below, the variance appears to be non-constant, which is also confirmed by the Shapiro-Wilk test on the residuals with a p-value of 0.0008193. Taking the square root or log of the response may help in resolving some of these issues.

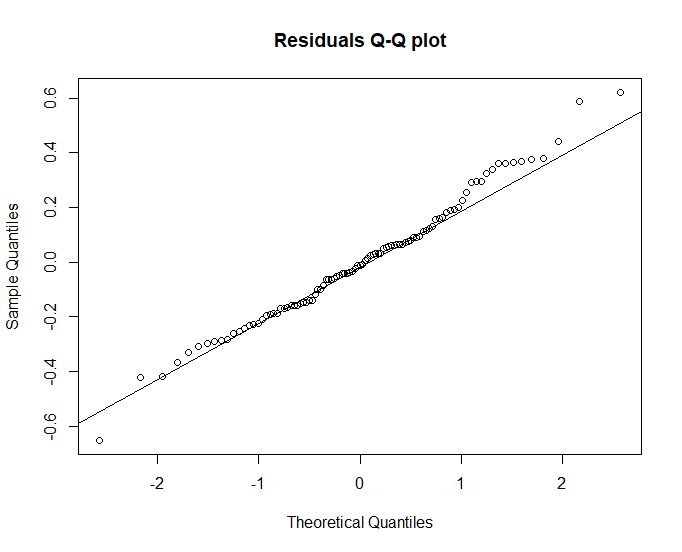


# Part 2

Transforming the data to take the logarithm of the response appears to fix the issues when looking at the residual vs fitted values plot. The data now appears to be normally distributed and have equal variance.



Performing a Shapiro-Wilk test on the residuals, we obtain a p-value of 0.519, meaning that we cannot reject the hypothesis that residuals are normally distributed. Looking at the Q-Q plot this appears to be the case as well. The right tail does appear to be slightly heavy and further actions may be taken to determine whether these points are valid.



CODE

> ## Import needed libraries and data

> library(reshape2)

> library(ggplot2)

> require(gridExtra)

>

> bear=read.table('bears.txt',header=TRUE,sep='\t')

> bear=bear[bear$Obs.No==1,]

>

> ## Linear System

> lm\_length = lm(bear$Weight~bear$Length)

> p1 <- ggplot(bear, aes(x=lm\_length$fitted.values, y=lm\_length$residuals)) + geom\_point(size=1) + geom\_hline(yintercept = 0, color="red")

> plot(p1)

> qqnorm(lm\_length$residuals,main='Residuals Q-Q plot')

> qqline(lm\_length$residuals)

> shapiro.test(lm\_length$residuals)

Shapiro-Wilk normality test

data: lm\_length$residuals

W = 0.94946, p-value = 0.0008193

>

> ## Log Transformed System

> lm\_log = lm(log(bear$Weight)~bear$Length)

> p2 <- ggplot(bear, aes(x=lm\_log$fitted.values, y=lm\_log$residuals)) + geom\_point(size=1) + geom\_hline(yintercept = 0, color="red")

> plot(p2)

> qqnorm(lm\_log$residuals,main='Residuals Q-Q plot')

> qqline(lm\_log$residuals)

> shapiro.test(lm\_log$residuals)

Shapiro-Wilk normality test

data: lm\_log$residuals

W = 0.98805, p-value = 0.519