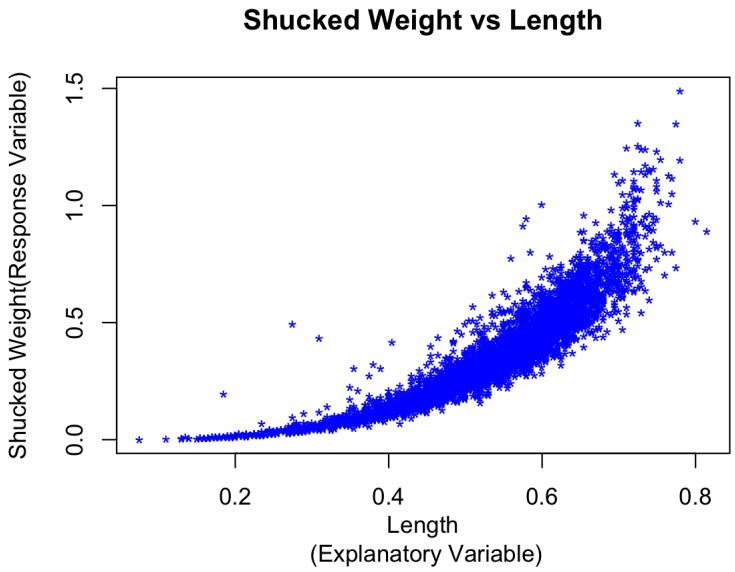
Amber Swain

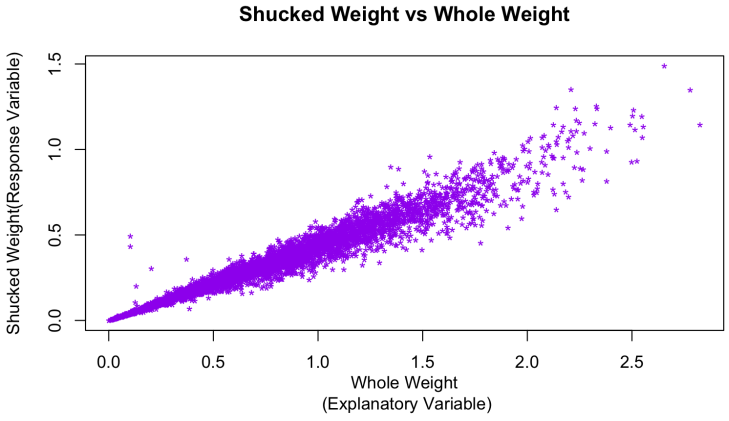
DSCI 407

Homework 1 Abalone Data Report

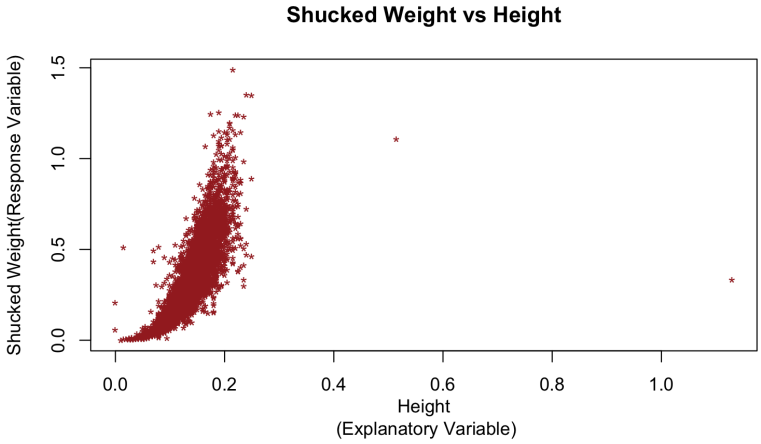
(a)



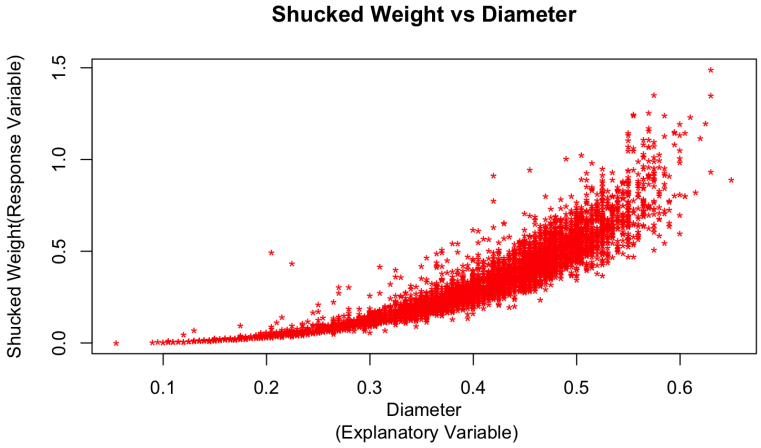
This shows the relationship between Length(Explanatory Variable) and Shucked Weight(Response Variable). The relationship appears to be non-linear because the line that is created is not a straight line and is more of an exponential or quadratic line. The strength is Weak, because as the length increased, the shucked weight did not increase at the same rate. This relationship has a strong positive correlation coefficient of 0.898. There seems to be at least a couple outliers at the points around (0.3,0.5) on the plot, and possibly at the point with the highest shucked weight.



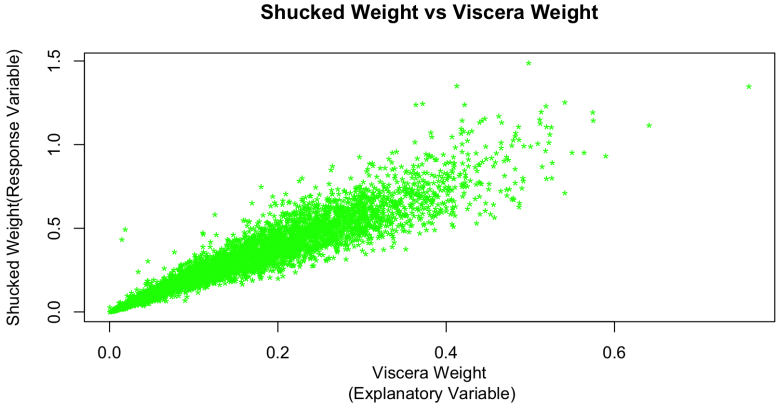
Here is the relationship between Shucked Weight and Whole Weight. It seems to be linear as it creates a mostly straight line, but has a handful of outliers around the point (0.1,0.5). The relationship has a strong positive correlation coefficient of 0.969 and its strength is strong.



This is the relationship between Shucked Weight and how the Height affects it. The relationship is non-linear as it exhibits more of an exponential curve but has a moderate-strong correlation coefficient of 0.775. There are obvious outliers on the right hand side of the graph and also possibly around the point (0.1,0.5). The strength is weak because each variable does not increase at the same or similar rate.

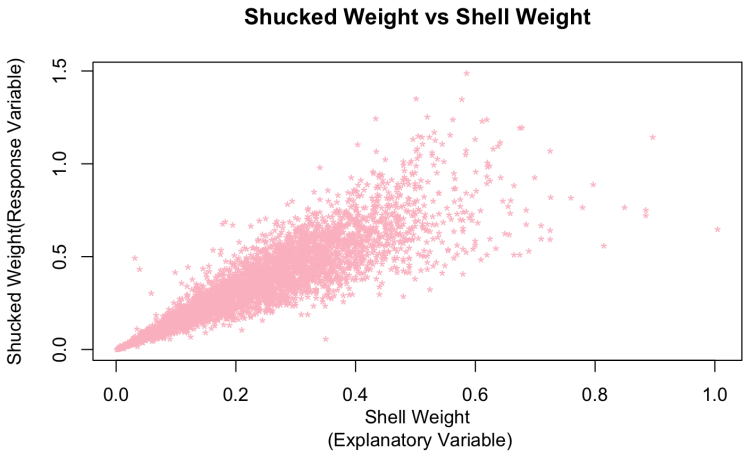


This is the relationship between Shucked Weight and the abalone's Diameter. It has a strong positive correlation coefficient of 0.893 yet exhibits a non-linear relationship, and more exponential or quadratic.The strength is weak and has some outliers in the upper right hand side of the plot.



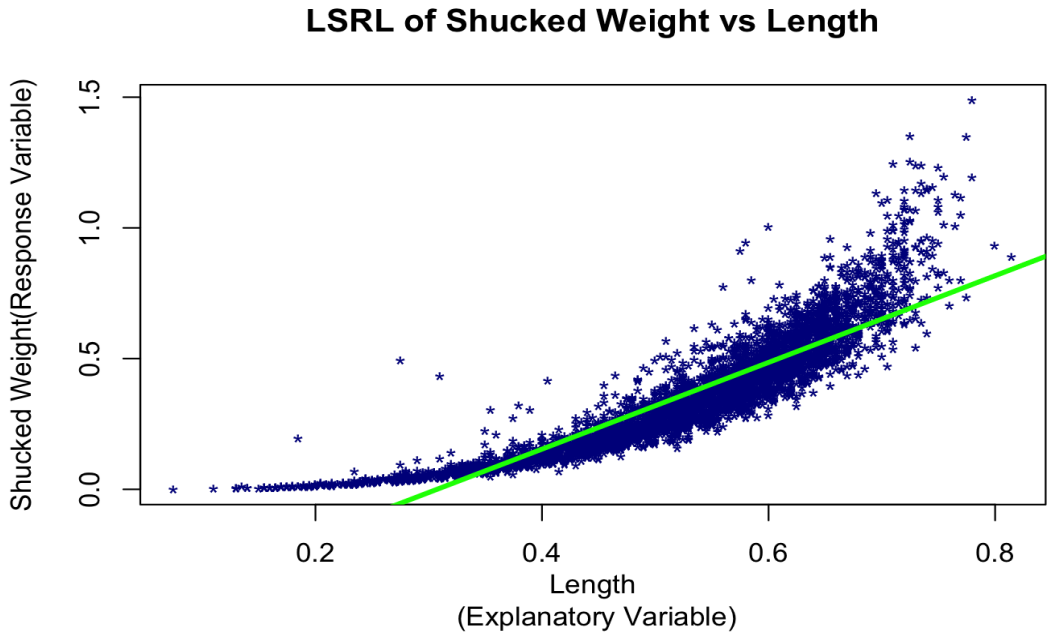
The Shucked Weight vs Viscera Weight has a linear relationship with a strong-positive correlation coefficient of 0.932. It has a strong strength and some outliers at the top right hand

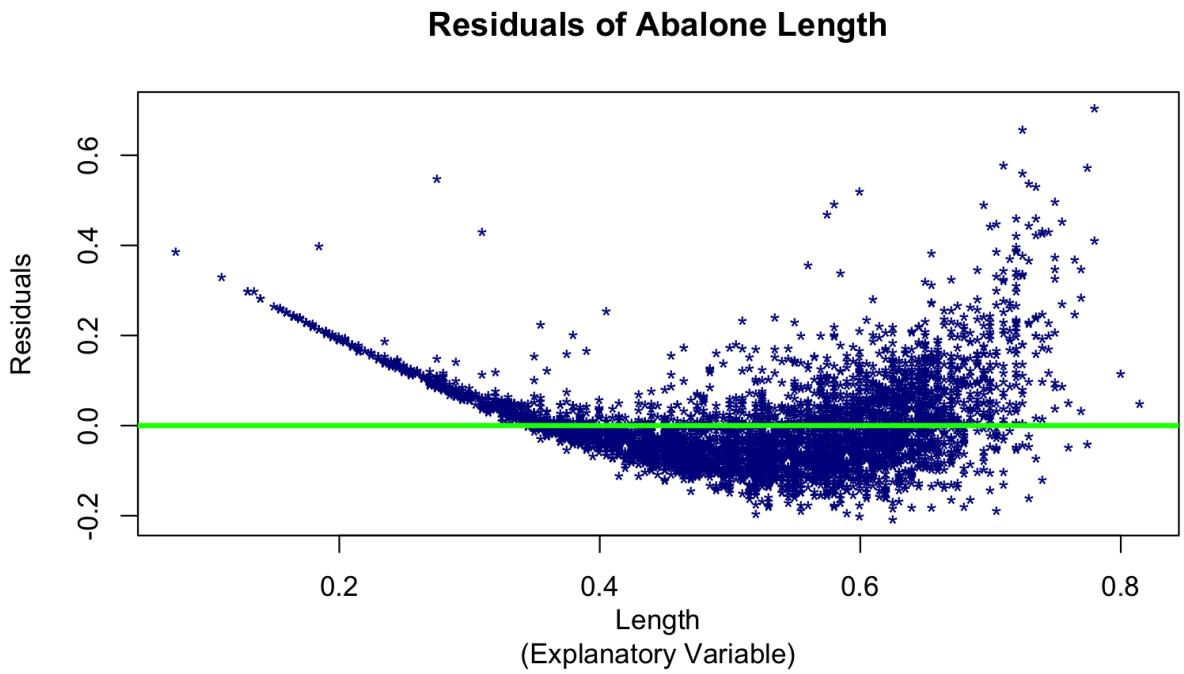
corner of the plot and around the point (0.1,0.5).

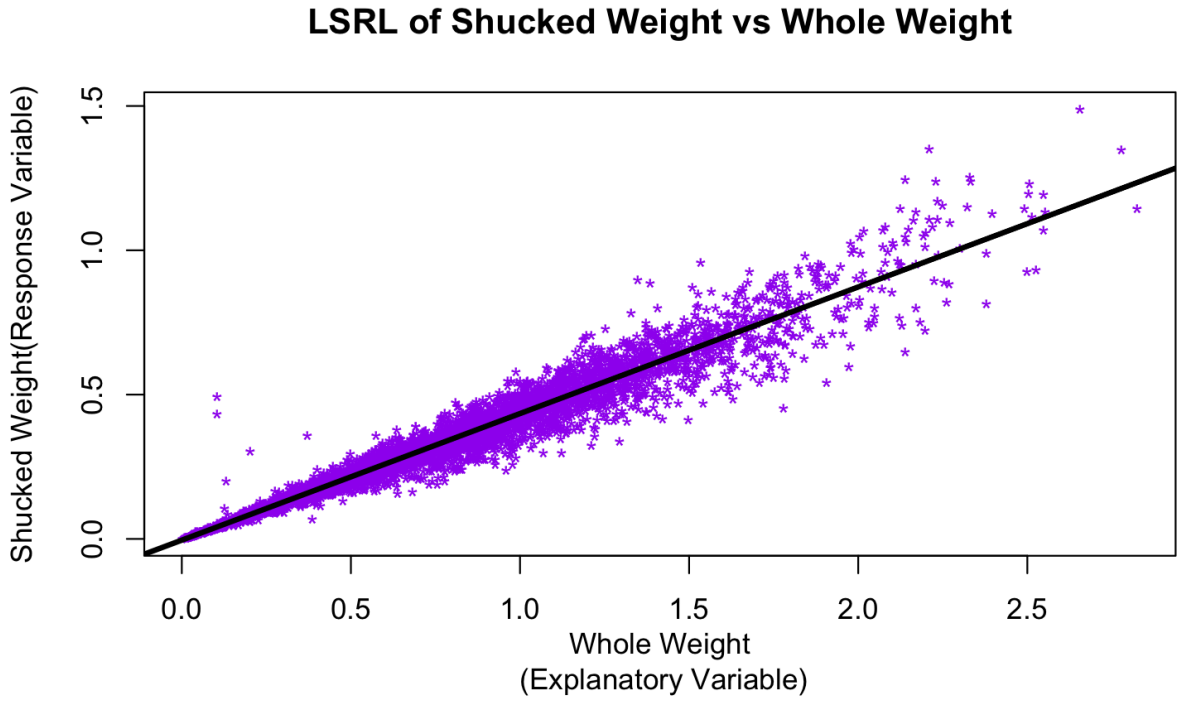


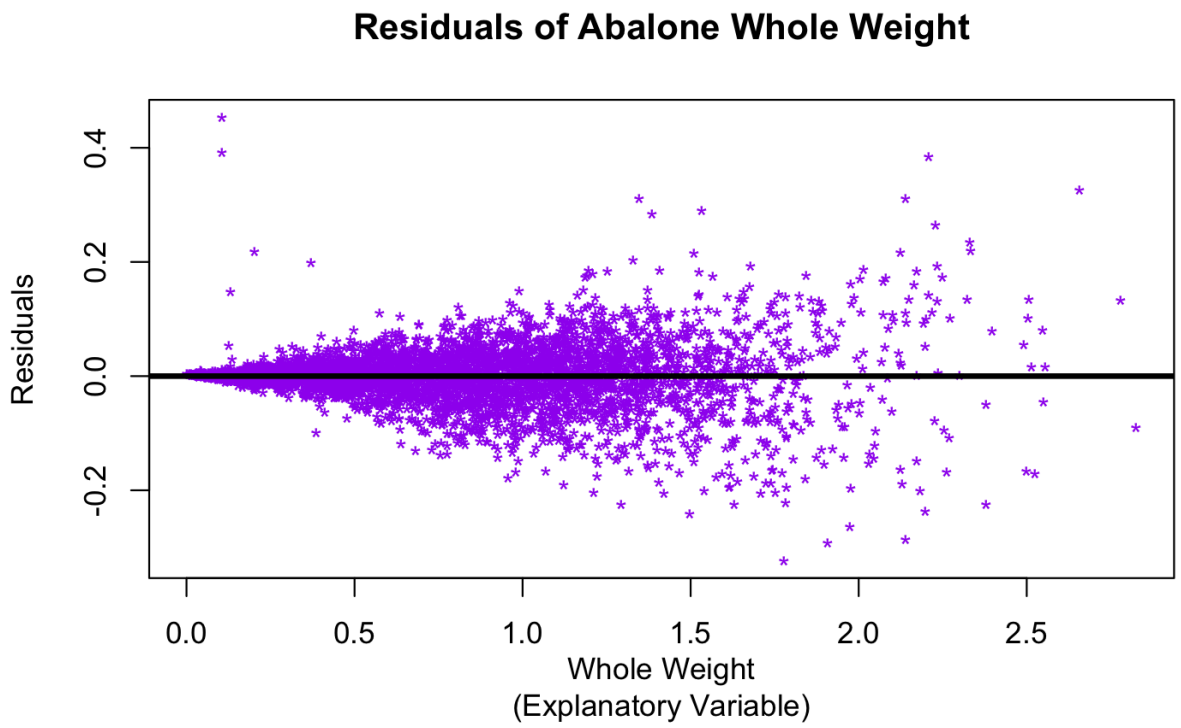
This is Shucked Weight vs Shell Weight, and it has a pretty similar relationship to Shucked Weight vs Viscera Weight. This relationship appears to be linear even though as the Shell Weight increases, the plots start to fan out. It has a moderate-strong correlation coefficient of 0.883 and has outliers on the right-middle side of the plot, around (0.6,1.5), and around (0.1,0.5). This plot also has a moderate strength.

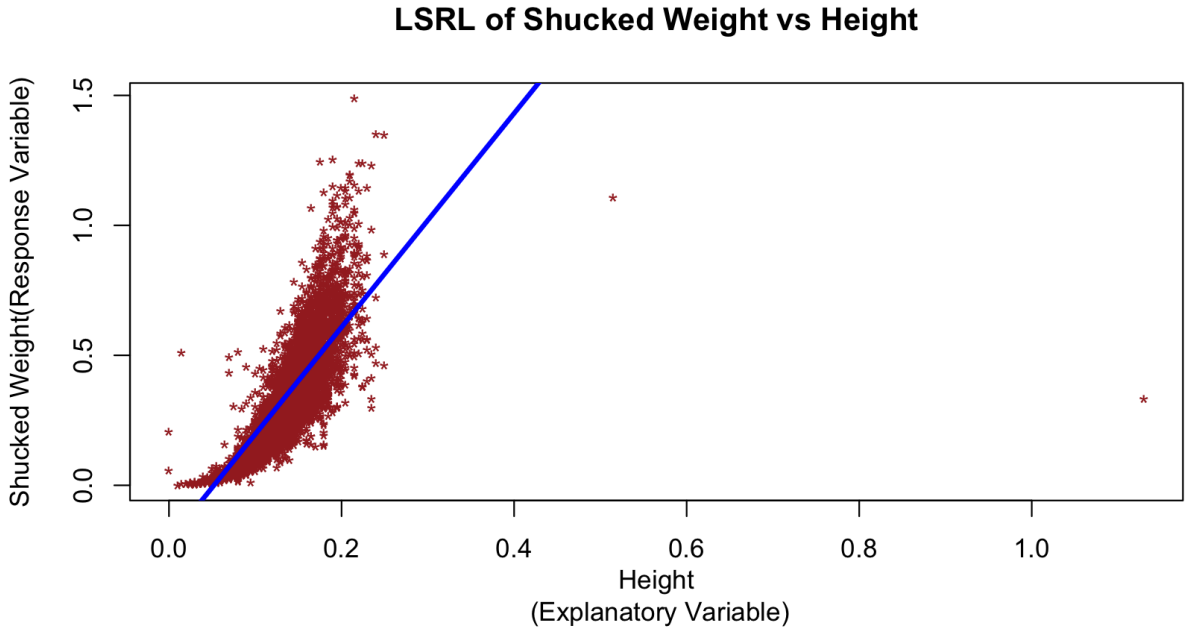
(b)

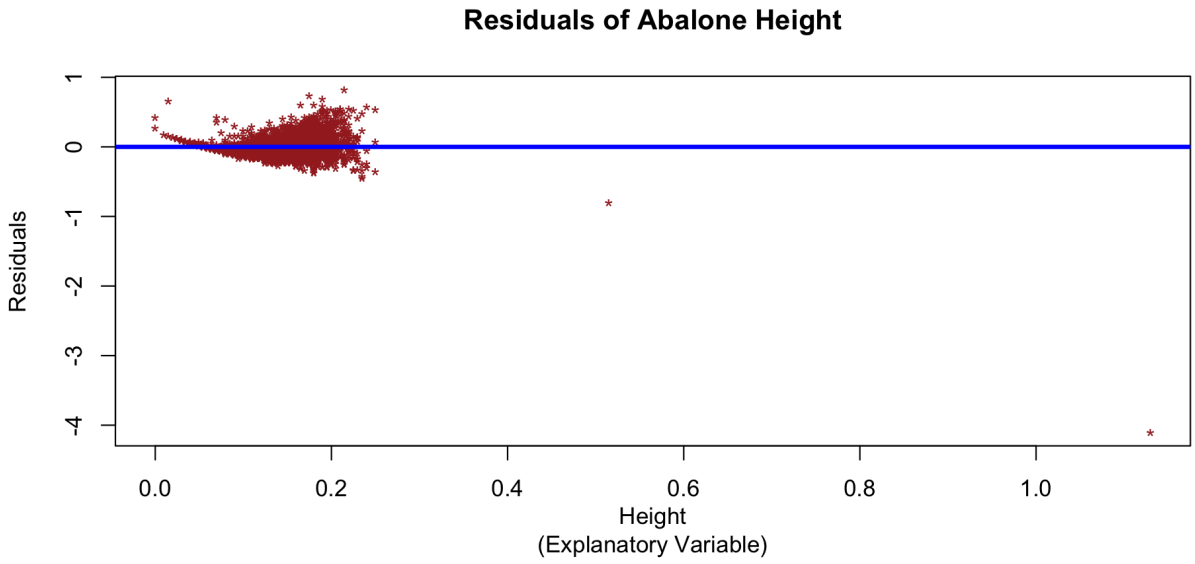


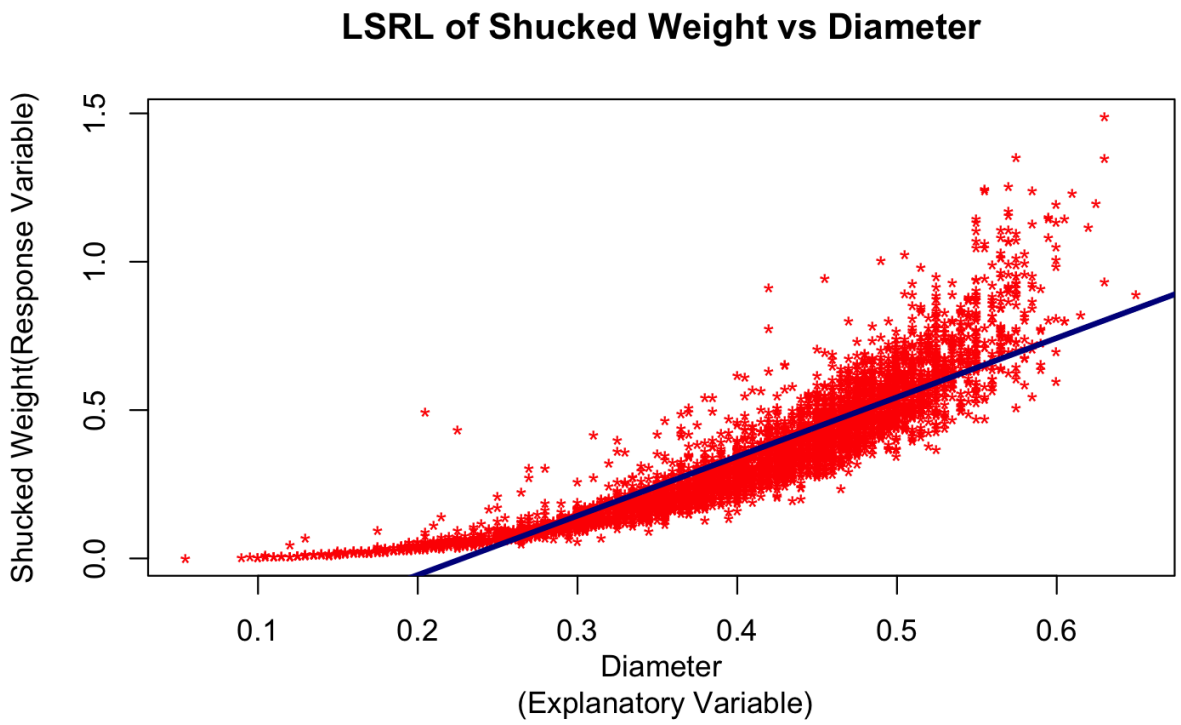

Residuals of Abalone Length
Residuals
-0.2
0.0
0.2
0.4
0.6
0.2
0.4
0.6
0.8
Length
(Explanatory Variable)

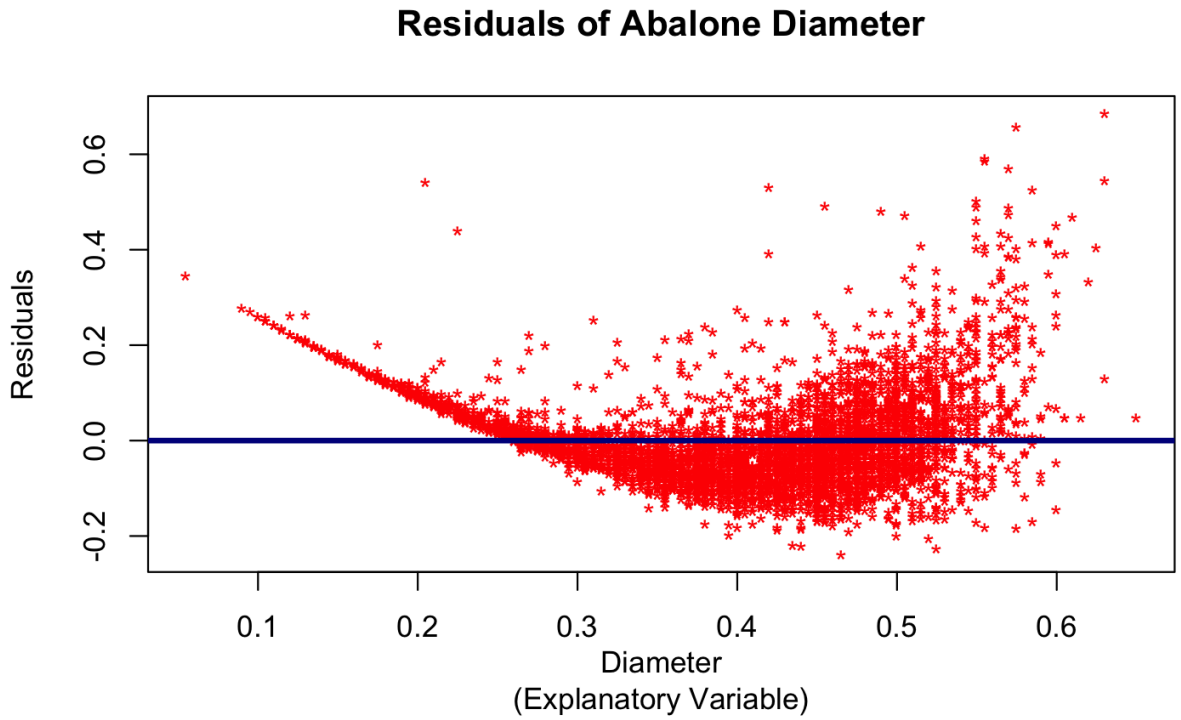

LSRL of Shucked Weight vs Whole Weight
Shucked Weight(Response Variable)
0.0
0.5
1.0
1.5
0.0
0.5
1.0
1.5
2.0
2.5
Whole Weight
(Explanatory Variable)

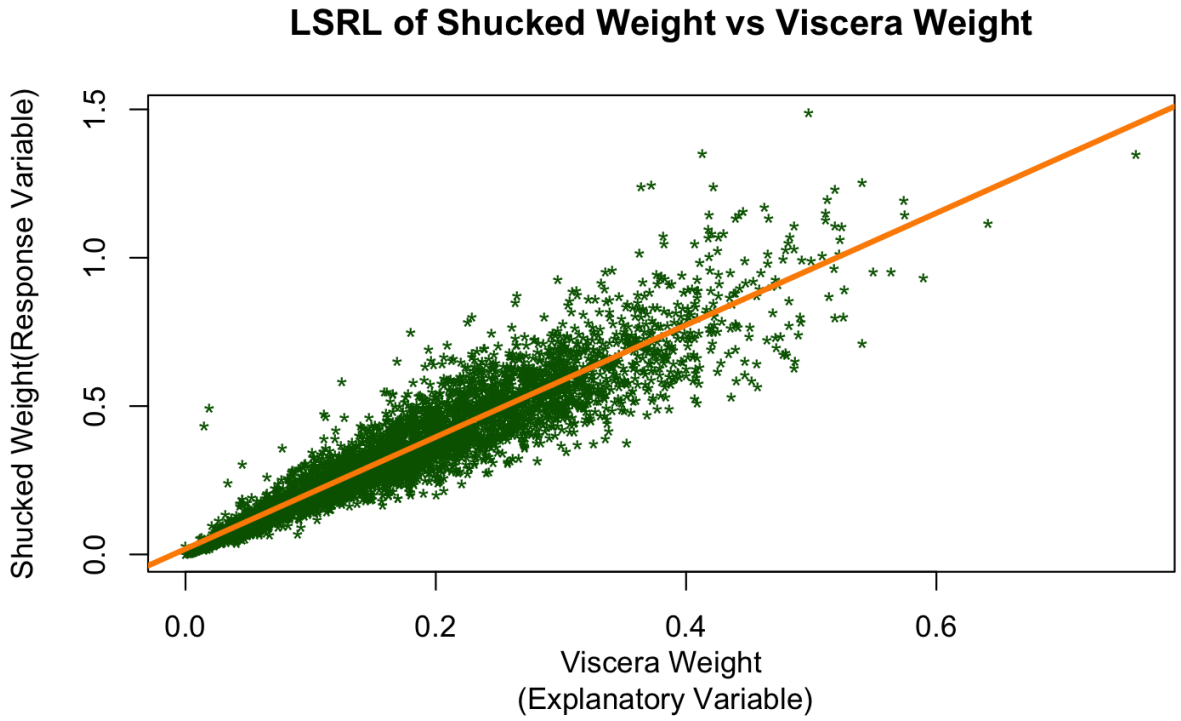

Residuals of Abalone Whole Weight
Residuals
0.4
0.2
0.0
-0.2
0.0
0.5
1.0
1.5
2.0
2.5
Whole Weight
(Explanatory Variable)

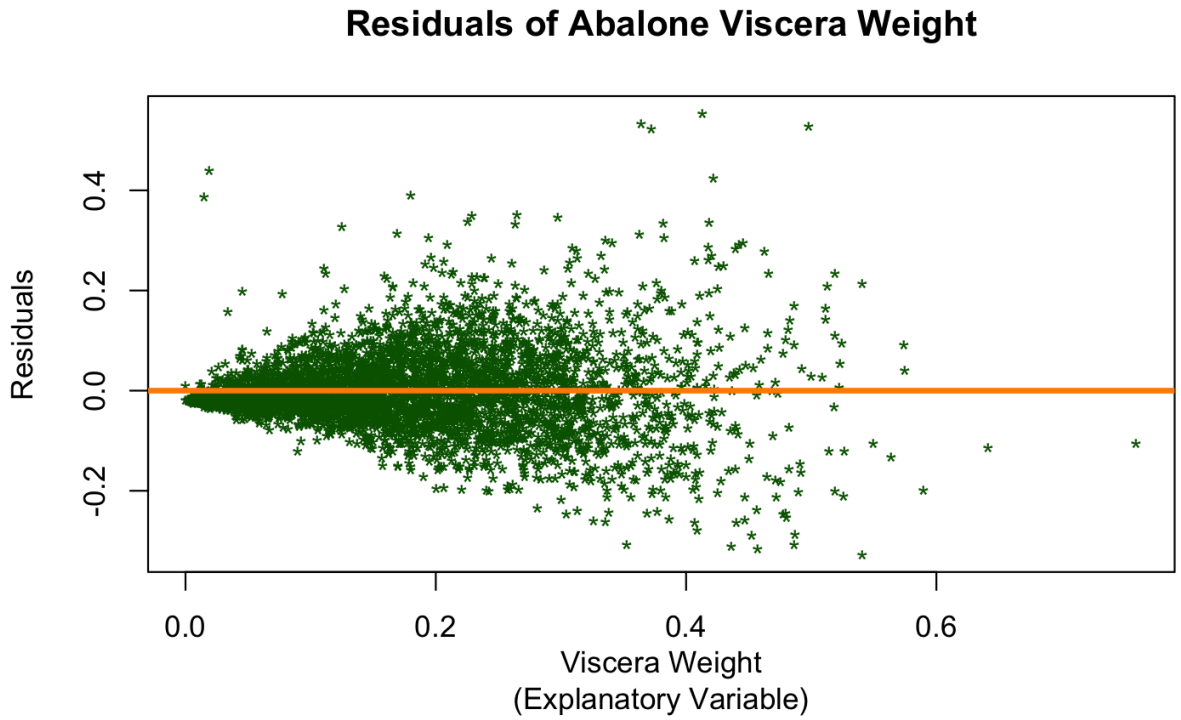

LSRL of Shucked Weight vs Height
Shucked Weight(Response Variable)
1.5
1.0
0.5
0.0
0.0
0.2
0.4
0.6
0.8
1.0
Height
(Explanatory Variable)

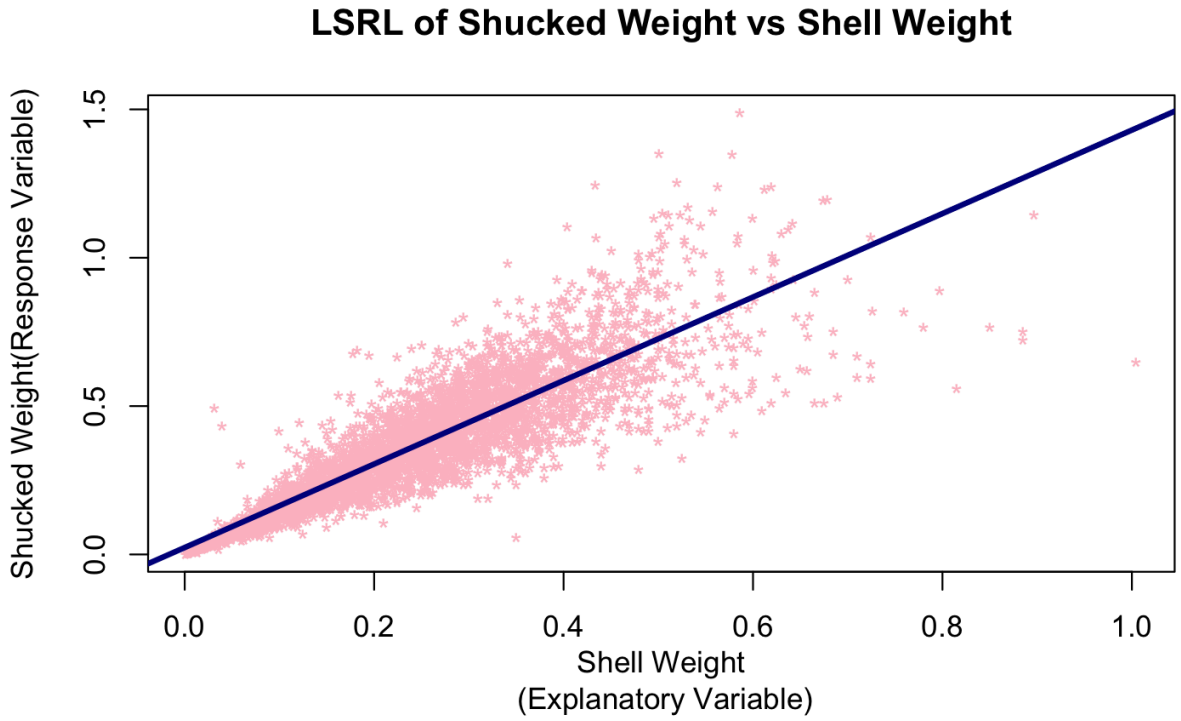

Residuals of Abalone Height
Residuals
1
0
-1
-2
-3
-4
0.0
0.2
0.4
0.6
0.8
1.0
Height
(Explanatory Variable)

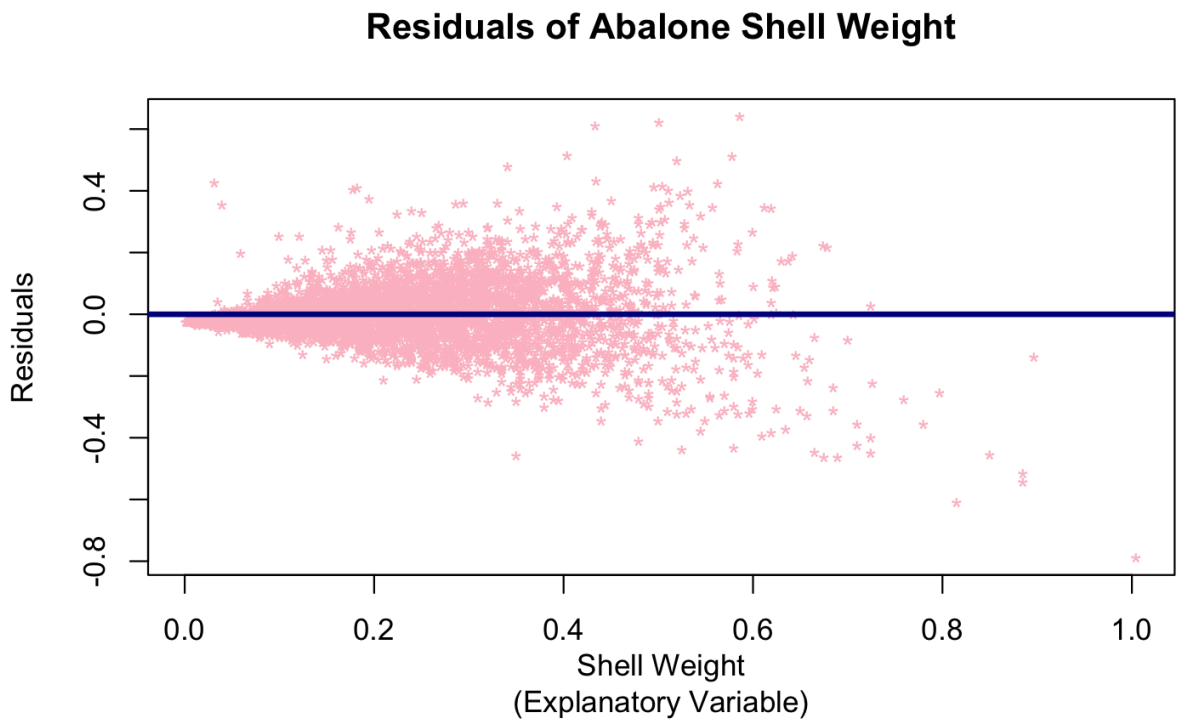

LSRL of Shucked Weight vs Diameter
Shucked Weight(Response Variable)
0.0
0.5
1.0
1.5
0.1
0.2
0.3
0.4
0.5
0.6
Diameter
(Explanatory Variable)


Residuals of Abalone Diameter
Residuals
-0.2
0.0
0.2
0.4
0.6
0.1
0.2
0.3
0.4
0.5
0.6
Diameter
(Explanatory Variable)


LSRL of Shucked Weight vs Viscera Weight
Shucked Weight(Response Variable)
0.0
0.5
1.0
1.5
0.0
0.2
0.4
0.6
Viscera Weight
(Explanatory Variable)


Residuals of Abalone Viscera Weight
Residuals
0.4
0.2
0.0
-0.2
0.0
0.2
0.4
0.6
Viscera Weight
(Explanatory Variable)

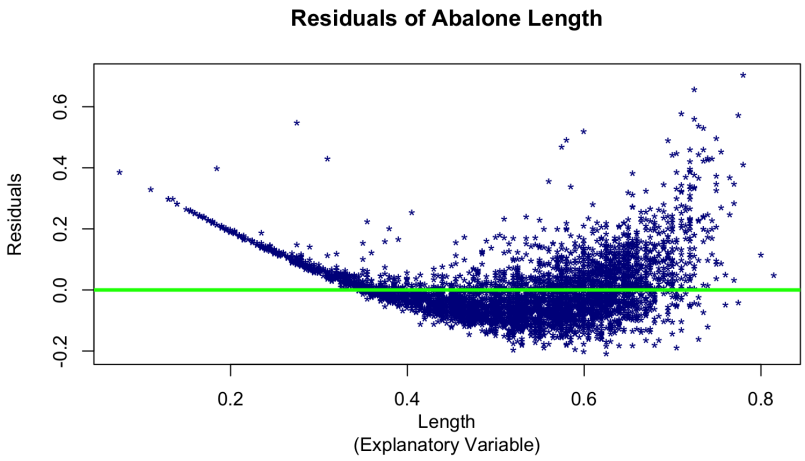

LSRL of Shucked Weight vs Shell Weight
Shucked Weight(Response Variable)
1.5
1.0
0.5
0.0
0.0
0.2
0.4
0.6
0.8
1.0
Shell Weight
(Explanatory Variable)



(c)

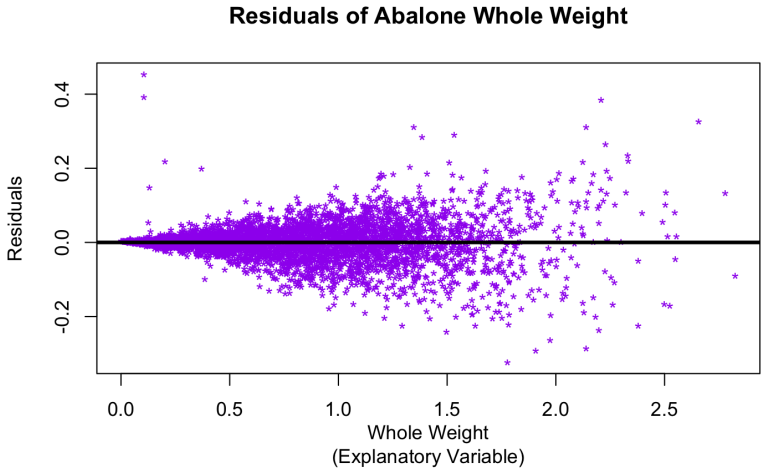
Conditions for Linear Model:

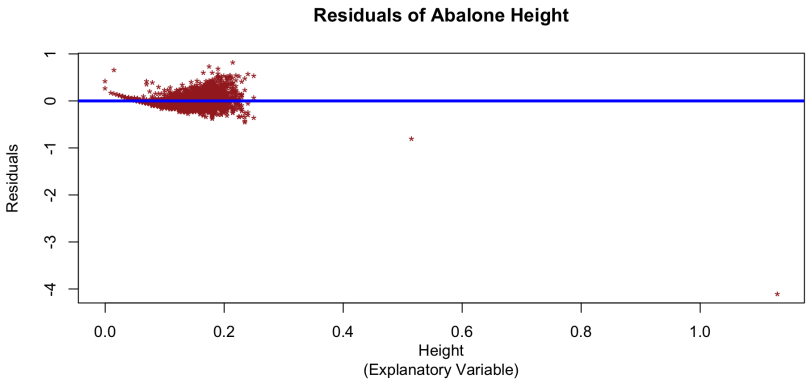
- No outliers
- No obvious pattern in Residual plot
- Check that the variance of the y-value is the same for all x-values
    - The variance is constant(Homoscedasticity)
    - No fanning out or funnels



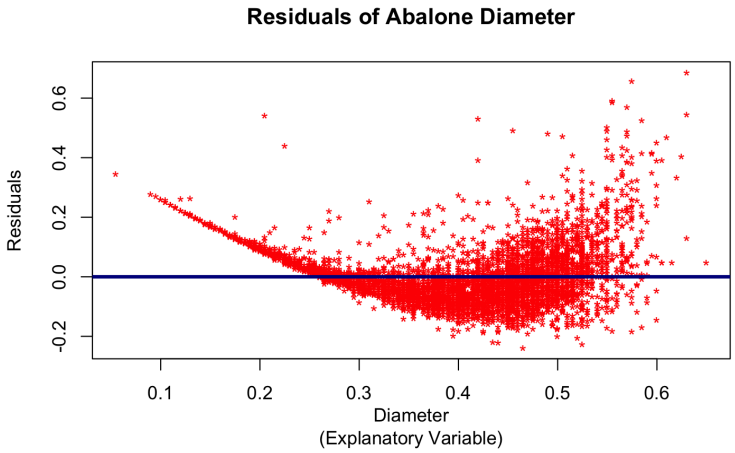
Assumptions are **not satisfied**, there is an obvious pattern(curve/exponential shape), there are many outliers and the variance is not constant.

Assumptions **are satisfied.**

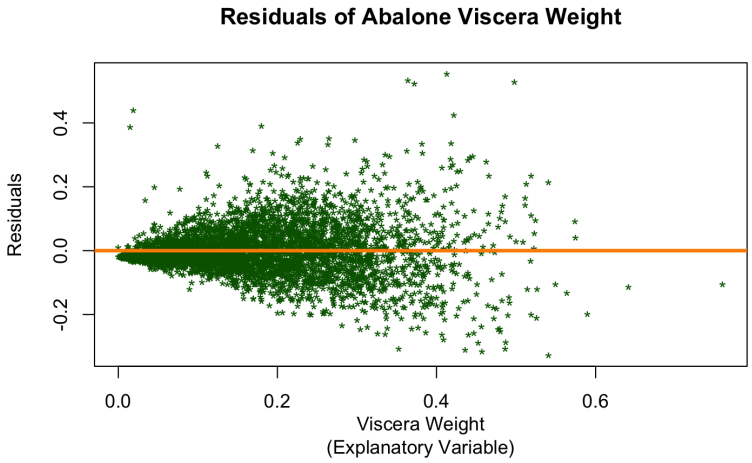




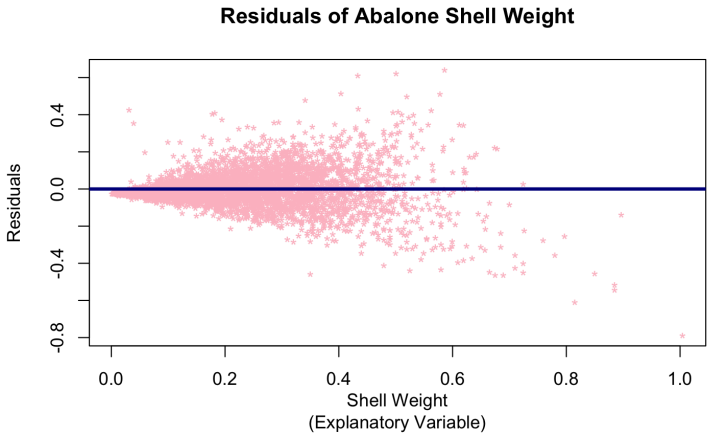
Assumptions are **not satisfied**; there are a couple drastic outliers and there is an obvious pattern in the plots(even without the outliers).



Assumptions are **not satisfied**; there is an obvious pattern in the residual plots, the data has some outliers, and the variance is not constant.



Assumptions are **not satisfied;** the plot has many outliers and the variance fans out drastically as the Viscera Weight increases.



Assumptions **are satisfied.**

(d)

<u>Whole Weight:</u>

- 95% Confidence Interval for Slope = [0.43540724 , 0.44214950]
- t-value = 255.179
- p-value = <2e-16***
- We have a significantly high t-value which in turn gives us a very low p-value. The three stars after the expression mean that the p-value is less than 0.001.
- Hypothesis Test
    - H0: B1 = 0

- H1: B1≠ 0
- Let α=0.05
- t = 255.179
- p = 2e-16
- Since our p-value is less than 0.05, we can reject our null hypothesis and accept the alternative hypothesis. This means that there is a significant association between an Abalone's Whole Wright(Explanatory Variable/Predictor) and its Shucked Weight(Response Variable/Outcome).

Shell Weight:

- 95% Confidence Interval for Slope = [1.38461636, 1.43010387]
- t-value = 121.316
- p-value = < 2e-16 ***
- Hypothesis Test
  - H0: B1 = 0
  - H1: B1≠ 0
  - Let α=0.05
  - t = 121.316
  - p = 2e-16
  - Since our p-value is less than 0.05, we can reject our null hypothesis and accept the alternative hypothesis. This means that there is a significant association between an Abalone's Shell Wright(Explanatory Variable/Predictor) and its Shucked Weight(Response Variable/Outcome).

(e)

Whole Weight: Prediction for when Whole Weight = 0.4415(Q1)

fit	lower	upper
0.189454	0.1873477	0.1915603

The "fit" is the predicted value of the Shucked Weight using the regression line. The lower and upper numbers are the bounds for the 95% confidence interval for the equivalent Shucked Weight.

Shell Weight:

fit	lower	upper
0.2062033	0.2021848	0.2102217

The “fit” is the predicted value of the Shucked Weight using the regression line. The lower and upper numbers are the bounds for the 95% confidence interval for the equivalent Shucked Weight.

(f)

I would say that you can use the Whole Weight to accurately predict your profit margin as an abalone farmer, solely because of how high the t-value was and in contrast how low the p-value was. The p-value was less than 0.001 which is a really low significance level which shows that there is a strong relationship between the two variables.