## Exercise 1

(a) To perform an analysis based on the given linear model on the dataset, one can run the following lines of code in the R console.

```
ModelLog <- lm(log(Mass) ~ Height + Width + Length, data = mussels)
summary(ModelLog)
```

From the summary, we obtain the estimates of the paramaters as follows

```
(Intercept)      Height       Width      Length
3.135116758 0.009421034 0.014325594 0.002362504
```

(b) We provide 95% confidence interval for our paramater $\beta_1$ as shown below.

```
> confint(ModelLog,'Height',level=0.95)
             2.5 %      97.5 %
Height 0.006766628 0.01207544
```

We can convince ourselves that $\beta_1 = 0$ lies in the rejected region, hence we choose to reject the hypothesis. This might suggests log(Mass) and Height are positively correlated to each other with $\beta_1 > 0$.

(c) To obtain the prediction on the specified input, we run the following code:

```
newdata<-data.frame(Height=100, Width=50, Length=200)
yhat<-predict(ModelLog,newdata)
yhat
```

This returns the value 5.266001.

(d) Now, we use again the predict function, with some additional arguments as shown below, to obtain the prediction with 95% confidence interval.

```
ypred<-predict(ModelLog,newdata,interval="prediction",level=0.95)
ypred
```

This gives us

```
       fit      lwr      upr
1 5.266001 5.099812 5.432189
```

(e) We run **plot(ModelLog)** to obtain the diagnostic plots of our linear model. The Scale-Location plot highlighted few potential outliers (points which standardised residuals above 1.5) that can be referred to the dataset. Plus, the Normal probability plot suggests a systematic error based on the deviation of the points at the tails, which might against our assumption on the distribution of the errors $\varepsilon_i$.

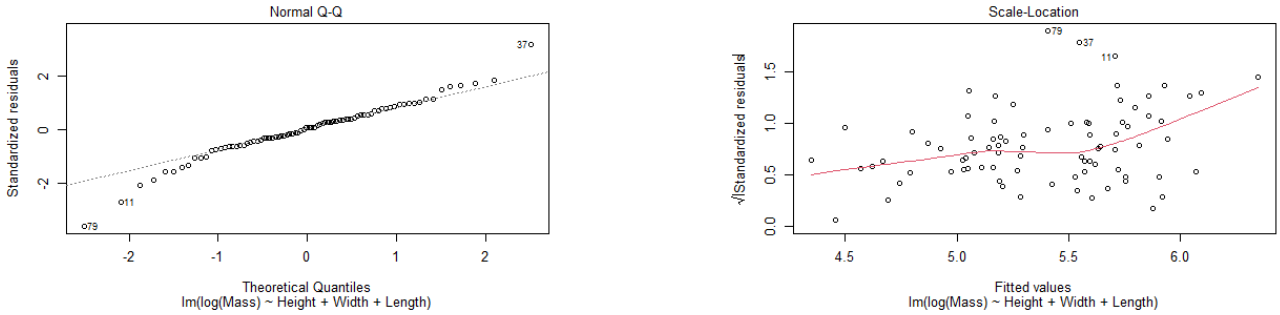


Figure 1: Two of the diagnostic plots of the model

```
   log(Mass) Height Width Length
11   5.52069    135    48    260
37   5.75737    129    46    228
79   5.15149    112    45    242
```

These values are very close to our `newdata` as defined in (c) hence might strongly influence our answers in (c) and (d)